

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2004, 12:22:52 ; Search time 55 Seconds  
(without alignments)  
35.961 Million cell updates/sec

Title: US-09-394-019C-212

Perfect score: 39

Sequence: 1 LEHGIN 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1017428

Minimum DB seq length: 0

Maximum DB seq length: 150

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	39	100.0	7	AAG73286	Protease
2	39	100.0	7	ABU60442	Protease
3	39	100.0	18	AAG73324	Protease
4	39	100.0	18	AAG73230	Protease
5	35	89.7	7	AAG73288	Protease
6	35	89.7	7	ABU60444	Protease
7	35	89.7	18	AAG73322	Protease
8	35	89.7	18	AAG73326	Protease
9	34	87.2	58	ABP01506	Human ORF
10	34	87.2	124	AU53518	Propionib
11	34	87.2	124	ABM50037	Propionib
12	31	79.5	30	AAY39352	Linker pe
13	30	76.9	7	AB27270	Murine CR
14	30	76.9	7	AAG73290	Protease
15	30	76.9	7	ABU60446	Protease
16	30	76.9	18	AAG73328	Protease
17	30	76.9	18	AAG73324	Protease
18	29	74.4	5	AAY15690	Peptide u
19	29	74.4	7	ABU60443	Protease
20	29	74.4	7	ABU60443	Protease
21	29	74.4	18	AAG73321	Protease
22	29	74.4	18	AAG73325	Protease
23	29	74.4	20	ABM50029	Babesia m
24	29	74.4	20	ADE06161	BMN1-re1a
25	29	74.4	38	AAG07964	Arabidops

## ALIGNMENTS

### RESULT 1

AAG73286

ID AAG73286 standard; peptide; 7 AA.

XX AAG73286;

XX 14-AUG-2001 (first entry)

XX DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

XX DE DT 14-AUG-2001 (first entry)

Aay55707 M. leprae  
Aaw60953 Streptococ  
Aag07962 Arabidops  
Aag49408 Arabidops  
Aar37963 KGFR Ig-1  
Aam84389 Human imm  
Aag49407 Arabidops  
Aag34647 Arabidops  
Aap80959 N. gonorr  
Ade95330 E. faeciu  
Aap64017 Human ORF  
Aag34646 Arabidops  
Aae32812 Human CEA  
Aae32812 Propionib  
Abm48521 Propionib  
Aau57868 Propionib  
Aag51770 Novel hum  
Abm54387 Propionib  
Abg27666 Novel hum  
Abg93329 C. albica

Protease indicator compound peptide #15.  
Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
haemophilia.  
Synthetic.  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US024882.  
10-SEP-1999; 99US-00394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the  
presence of active proteases, useful for detecting and localizing  
protease activity in biological samples, particularly in frozen tissue  
samples.  
Claim 1; Page 70; 86pp; English.  
The present invention describes fluorogenic compositions which can be  
used for the detection of protease activity. This can be useful as an  
indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
thrombosis and arthritis. The fluorogenic compositions comprise a  
peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
peptide is cleaved by a protease and the fluorophores can then be  
detected. The present sequence is one of the peptides described in the  
exemplification of the invention  
Sequence 7 AA;

Query Match 100.0%; Score 39; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
 |||||  
 Db 1 LEHDGIN 7

## RESULT 2

ID ABU60442  
 ABU60442 standard; peptide; 7 AA.

XX AC ABU60442;

DT 29-APR-2003 (first entry)

XX Protease binding peptide motif SEQ ID 166.

XX Protease; indicator; chromophore; H-dimer; fluorescence; absorbance;  
 KW nuclease; screening; fluorophore; substrate cleavage.

XX Synthetic.

XX WO200261038-A2.

XX 08-AUG-2002.

XX 21-DEC-2001; 2001WO-US049781.

XX 22-DEC-2000; 2000US-00747287.

XX (ONCO-) ONCOIMMUNIN INC.

XX Packard BS, Komoriya A;

XX WPI; 2002-698548/75.

PT Indicator composition comprising polypeptide or nucleic acid backbone  
 joining two same chromophores resulting in quenching of fluorescence  
 of change in absorbance of chromophores, useful for detecting protease  
 activity.

PS Disclosure; Page 35; 97pp; English.

CC This invention describes a novel indicator composition (referred as homo-  
 CC doubly labeled compositions) comprising a polypeptide backbone or a  
 CC nucleic acid backbone joining two chromophores of the same species  
 CC whereby the chromophores form an H-dimer resulting in quenching of the  
 CC fluorescence of or a change in the absorbance of the chromophore, a  
 CC decrease in fluorescence or a change in absorbance indicates that the  
 CC first molecule and the second molecule are interacting. The indicator is  
 CC useful for detecting the activity of a protease, where an increase in  
 CC fluorescence or a change in absorbance indicates that the protease  
 CC cleaves the polypeptide backbone. The indicator is attached to a solid  
 CC support inside a mammalian, yeast or insect cell. The composition bears a  
 CC hydrophobic group such as Fmoc, 9-fluorenylmethyl group, 1-  
 CC fluoronaphthylmethyl group, 9-fluorenylmethyl group, and 9-fluorenylmethyl-  
 CC carboxylic group, benzoyloxycarbonyl, Xanthyl (Xan), Trityl (Ttt), 4-  
 CC methyltrityl (Mtt), 4-methoxytrityl (Mmt), 4-methoxy-2,3,6-trimethyl-  
 CC benzenehexyl (Mtr), Mesitylene-2-sulphonyl (Mts), 4,4'-  
 CC dimethoxybenzyl (Mdb), etc. The method described in the invention is  
 CC useful for detecting protease or nuclease activity (or the presence of  
 CC nucleic acid) in histological section, cells in culture, (e.g., seeded or  
 CC cultured adherent cells), a biological sample such as tissue, biopsy,  
 CC lymph, embryo, or whole animal, or cell suspension derived from a  
 CC biological sample such as tissue, blood, urine, saliva, lymph, or biopsy.  
 CC The indicator composition is also useful for screening a test agent for  
 CC the ability to modulate a protease (or a nuclease, lipase, etc.). The  
 CC indicator reagents allow rapid determination of protease activity in a  
 CC matter of minutes in a single-step procedure. The fluorescent indicators  
 CC both absorb and emit in the visible range (400-800 nm). These signals are  
 CC therefore not readily quenched by, nor is activation of the fluorophores,

CC that is, absorption of light, interfered with by background molecules;  
 CC therefore they are easily detected in biological samples. The fluorogenic  
 CC protease indicators utilize high efficiency fluorophores and are able to  
 CC achieve a high degree of quenching while providing a strong signal when  
 CC the quench is released by cleavage of the peptide substrate. The high  
 CC signal allows detection of very low levels of protease activity. Thus the  
 CC fluorogenic protease indicators are particularly well suited for in situ  
 CC detection of protease activity. ABU60357-ABU60477 represent peptides use  
 CC to illustrate the method described in the disclosure of the invention

XX Sequence 7 AA;

Query Match 100.0%; Score 39; DB 5; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
 |||||  
 Db 1 LEHDGIN 7

## RESULT 3

AAG73324

ID AAG73324 standard; peptide; 18 AA.

XX AC AAG73324;

XX 14-AUG-2001 (first entry)

XX Protease indicator compound peptide #53.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
 KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
 KW haemophilia.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 4  
 /label= OTHER  
 /note= "designated J in the specification"

FT Modified-site 14  
 /label= OTHER  
 /note= "designated J in the specification"

XX WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US024882.

XX 10-SEP-1999; 99US-00394019.

XX (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

XX New fluorogenic compositions whose fluorescence level increases in the  
 PT presence of active proteases, useful for detecting and localizing  
 PT protease activity in biological samples, particularly in frozen tissue  
 PT samples.

XX Claim 4; Page 71; 86pp; English.

XX The present invention describes fluorogenic compositions which can be  
 CC used for the detection of protease activity. This can be useful as an  
 CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
 CC thrombosis and arthritis. The fluorogenic compositions comprise a  
 CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
 CC peptide is cleaved by a protease and the fluorophores can then be  
 CC detected. The present sequence is one of the peptides described in the

CC exemplification of the invention  
XX  
SQ Sequence 18 AA;  
  
Query Match 100.0%; Score 39; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 LEHDGIN 7  
Db 6 LEHDGIN 12  
  
RESULT 4  
AAG73230  
ID AAG73230 standard; peptide; 18 AA.  
XX  
AC AAG73230;  
XX  
DT 14-AUG-2001 (first entry)  
XX  
DE Protease binding site #164.  
XX  
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
KW haemophilia.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"  
XX  
PN WO200118238-A1.  
XX  
PD 15-MAR-2001.  
XX  
PF 11-SEP-2000; 2000WO-US024882.  
XX  
PR 10-SEP-1999; 99US-00394019.  
XX  
PA (ONCO-) ONCOIMMUNIN INC.  
XX  
PI Komoriya A, Packard BS;  
XX  
DR WPI; 2001-389573/41.  
XX  
CC New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples.  
PS Disclosure; Page 29; 86pp; English.  
XX  
XX The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention  
XX  
SQ Sequence 18 AA;  
  
Query Match 100.0%; Score 39; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 LEHDGIN 7  
Db 6 LEHDGIN 12  
  
RESULT 5  
AAG73288  
ID AAG73288 standard; peptide; 7 AA.  
XX  
AC AAG73288;  
XX  
DT 14-AUG-2001 (first entry)  
XX  
DE Protease indicator compound peptide #17.  
XX  
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
KW haemophilia.  
XX  
OS Synthetic.  
XX  
PN WO200118238-A1.  
XX  
PD 15-MAR-2001.  
XX  
PF 11-SEP-2000; 2000WO-US024882.  
XX  
PR 10-SEP-1999; 99US-00394019.  
XX  
PA (ONCO-) ONCOIMMUNIN INC.  
XX  
PI Komoriya A, Packard BS;  
XX  
DR WPI; 2001-389573/41.  
XX  
CC New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples.  
PS Claim 1; Page 70; 86pp; English.  
XX  
XX The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention  
XX  
SQ Sequence 7 AA;  
  
Query Match 89.7%; Score 35; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 EHDGIN 7  
Db 2 EHDGIN 7  
  
RESULT 6  
ABU60444  
ID ABU60444 standard; peptide; 7 AA.  
XX  
AC ABU60444;  
XX  
DT 29-APR-2003 (first entry)  
XX  
DE Protease binding peptide motif SEQ ID 168.  
XX  
XX Protease; indicator; chromophore; H-dimer; fluorescence; absorbance;  
KW nuclease; screening; fluorophore; substrate cleavage.  
XX  
OS Synthetic.

WO200261038-A2.  
08-AUG-2002.  
21-DEC-2001; 2001WO-US049781.  
22-DEC-2000; 2000US-00747287.  
(ONCO-) ONCOIMMUNIN INC.  
Packard BS, Komoriya A;  
WPI; 2002-698548/75.  
Indicator composition comprising polypeptide or nucleic acid backbone joining two same chromophores resulting in quenching of fluorescence or change in absorbance of chromophores, useful for detecting protease activity.  
Disclosure; Page 35; 97pp; English.  
This invention describes a novel indicator composition (referred as homodoubly labeled compositions) comprising a polypeptide backbone or a nucleic acid backbone joining two chromophores of the same species whereby the chromophores form an H-dimer resulting in quenching of the fluorescence or a change in the absorbance of the chromophore, a decrease in fluorescence or a change in absorbance indicates that the first molecule and the second molecule are interacting. The indicator is useful for detecting the activity of a protease, where an increase in fluorescence or a change in absorbance indicates that the protease cleaves the polypeptide backbone. The indicator is attached to a solid support inside a mammalian, yeast or insect cell. The composition bears a hydrophobic group such as Fmoc, 9-fluoreneacetyl group, 1-carboxy-9-fluoreneacetyl group, 9-fluoreneacetyl group, and 9-fluorenone-1-carboxylic group, benzoyloxycarbonyl, Xanthyl (Xan), Trityl (Trt), 4-methyltrityl (Mtt), 4-methoxytrityl (Mmt), 4-methoxy-2,3, 6-trimethylbenzenesulfonyl (Mtr), Mesitylene-2-sulphonyl (Mts), 4,4'-dimethoxybenzhydryl (Mbh), etc. The method described in the invention is useful for detecting protease or nuclease activity (or the presence of nucleic acid) in histological section, cells in culture, (e.g., seeded or cultured adherent cells), a biological sample such as tissue, biopsy, lymph, embryo, or whole animal, or cell suspension derived from a biological sample such as tissue, blood, urine, saliva, lymph, or biopsy. The indicator composition is also useful for screening a test agent for the ability to modulate a protease (or a nuclease, lipase, etc.). The indicator reagents allow rapid determination of protease activity in a matter of minutes in a single-step procedure. The fluorescent indicators both absorb and emit in the visible range (400-800 nm). These signals are therefore not readily quenched by, nor is activation of the fluorophores, that is, absorption of light, interfered with by background molecules; therefore they are easily detected in biological samples. The fluorogenic protease indicators utilize high efficiency fluorophores and are able to achieve a high degree of quenching while providing a strong signal when the quench is released by cleavage of the peptide substrate. The high signal allows detection of very low levels of protease activity. Thus the fluorogenic protease indicators are particularly well suited for in situ detection of protease activity. AB060337-AB060477 represent peptides used to illustrate the method described in the disclosure of the invention

```

XX:      Sequence 7 AA;
SQ
      Query Match      89.7%;      Score 35;      DB 5;      Length 7;
      Best Local Similarity 100.0%;      Pred. NO. 1.4e+06;
      Matches 6;      Conservative 0;      Mismatches 0;      Indels 0;      Gaps 0;
      QY      2      EHDGIN 7
      Db      2      EHDGIN 7

```

RESULT 7  
AAG73232

ID	XX	AAG73232 standard; peptide; 18 AA.
XX	XX	
AC	AC	AAG73232;
XX	XX	
DT	DT	14-AUG-2001 (first entry)
XX	XX	
DE	DE	Protease binding site #166.
XX	XX	
XX	XX	Protease detection; peptide cleavage; enzyme activity; fluorogenic;
KW	KW	viral infection; cancer metastasis; emphysema; arthritis; thrombosis;
KW	KW	haemophilia.
XX	XX	
OS	OS	Synthetic.
XX	XX	
FH	FH	Key Location/Qualifiers
FT	FT	Modified-site 1
FT	FT	/label= OTHER
FT	FT	/note= "modified by fluorophore"
XX	XX	
PN	PN	WO200118238-A1.
XX	XX	
XX	XX	15-MAR-2001.
PD	PD	
XX	XX	
FF	FF	11-SEP-2000; 2000WO-US024882.
XX	XX	
PR	PR	10-SEP-1999; 99US-00394019.
XX	XX	
PA	PA	(ONCO-) ONCOIMMUNIN INC.
XX	XX	
FI	FI	Komoriya A, Packard BS;
XX	XX	
DR	DR	WPI; 2001-389573/41.
XX	XX	
PT	PT	New fluorogenic compositions whose fluorescence level increases in the
PT	PT	presence of active proteases, useful for detecting and localizing
PT	PT	protease activity in biological samples, particularly in frozen tissue
XX	XX	samples.
XX	XX	
PS	PS	Disclosure; Page 29; 86pp; English.
XX	XX	
CC	CC	The present invention describes fluorogenic compositions which can be
CC	CC	used for the detection of protease activity. This can be useful as an
CC	CC	indicator of viral infection, cancer metastasis, haemophilia, emphysema,
CC	CC	thrombosis and arthritis. The fluorogenic compositions comprise a
CC	CC	peptide, a peptide spacer and a donor and an acceptor fluorophore. The
CC	CC	peptide is cleaved by a protease and the fluorophores can then be
CC	CC	detected. The present sequence is one of the peptides described in the
CC	CC	exemplification of the invention
XX	XX	
SQ	SQ	Sequence 18 AA;
		Query Match 89.7%; Score 35; DS 4; Length 18;
		Best Local Similarity 100.0%; Pred. No. 3;
		Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	QY	2 EHDGIN 7
DB	DB	7 EHDGIN 12
RESULT 8		
AAG73326		
ID	XX	AAG73326 standard; peptide; 18 AA.
XX	XX	
AC	AC	AAG73326;
XX	XX	
DT	DT	14-AUG-2001 (first entry)
XX	XX	
XX	XX	Protease indicator compound peptide #55.
XX	XX	
XX	XX	Protease detection; peptide cleavage; enzyme activity; fluorogenic;
KW	KW	viral infection; cancer metastasis; emphysema; arthritis; thrombosis;
KW	KW	haemophilia.



XX OS Synthetic.

XX FH Key

XX FT Location/Qualifiers

XX FT 4

XX FT Modified-site

XX FT /label= OTHER

XX FT /note= "designated J in the specification"

XX FT 14

XX FT Modified-site

XX FT /label= OTHER

XX FT /note= "designated J in the specification"

XX FT

XX PN WO200118238-A1.

XX PD

XX PD 15-MAR-2001.

XX PF

XX PF 11-SEP-2000; 2000WO-US024882.

XX PR

XX PR 10-SEP-1999; 99US-00394019.

XX PR (ONCO-) ONCOIMMUNIN INC.

XX PA

XX PI Komoriya A, Packard BS;

XX PI WPI; 2001-389573/41.

XX DR

XX DR New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples.

XX PT

XX PS Claim 4; Page 71; 86pp; English.

XX CC The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention

XX SQ Sequence 18 AA;

Query Match 89.7%; Score 35; DB 4; Length 18;

Best Local Similarity 100.0%; Pred. No. 3;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7

Db 7 EHDGIN 12

RESULT 9

ABP01506

ID ABP01506 standard; protein; 58 AA.

XX AC ABP01506;

XX AC

XX DT 24-JUN-2002 (first entry)

XX DE Human ORFX protein sequence SEQ ID NO:2994.

XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis; hyperproliferative disorder; psoriasis; benign tumour; haemorrhage; degenerative disorder; osteoarthritis; neurodegenerative disorder; cardiovascular disease; diabetes mellitus; systemic lupus erythematosus; hypertension; hypothyroidism; cholesterol ester storage disease; immune deficiency; immune disorder; infectious disease; autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis; myasthenia gravis.

XX KW Homo sapiens.

OS XX

PN WO200192523-A2.

XX PD

XX PD 06-DEC-2001.

XX PF

XX PF 29-MAY-2001; 2001WO-US010836.

XX PR

XX PR 30-MAY-2000; 2000US-0206132P.

XX PR 29-AUG-2000; 2000US-0228716P.

XX PR (CURA-) CURAGEN CORP.

XX PA

XX PI Shimkets RA, Leach MD;

XX PI WPI; 2002-106308/14.

XX DR

XX DR N-PSDB; ABN17258.

XX PT

XX PT Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders.

XX PS Disclosure; SEQ ID NO 2994; 1037pp; English.

XX CC The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 58 AA;

Query Match 87.3%; Score 34; DB 5; Length 58;

Best Local Similarity 85.7%; Pred. No. 17;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGIN 7

Db 16 LEHDGIS 22

RESULT 10

AAU53518

ID AAU53518 standard; protein; 124 AA.

XX AC AAU53518;

XX AC

XX DT 27-FEB-2002 (first entry)

XX DE Propionibacterium acnes immunogenic protein #1414.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis; uveitis; endophthalmitis; bone; joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.

XX KW

OS Propionibacterium acnes.  
XX WO200181581-A2.  
XX  
XX  
PD 01-NOV-2001.  
XX  
XX 20-APR-2001; 2001WO-US012865.  
XX  
XX 21-APR-2000; 2000US-0199047P.  
XX PR 02-JUN-2000; 2000US-0208841P.  
XX PR 07-JUL-2000; 2000US-0216747P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
XX WPI; 2001-616774/71.  
XX DR N-PSDB; AAS95560.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX  
XX Example 1; SEQ ID NO 14713; 1069pp; English.  
XX  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 124 AA;  
SQ  
Query Match 87.2%; Score 34; DB 4; Length 124;  
Best Local Similarity 85.7%; Pred. No. 39;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LEHDGIN 7  
Db 34 LEHDGIS 40  
| | | | |  
RESULT 11  
ABM50037  
ID ABM50037 standard; protein; 124 AA.  
XX  
XX ABM50037;  
XX  
XX 20-OCT-2003 (first entry)  
XX  
XX Propionibacterium acnes predicted ORF-encoded polypeptide #14713.  
XX  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
XX immunostimulant; immune response; vaccine.  
XX  
XX Propionibacterium acnes.  
XX

PN WO2003033515-A1.  
XX  
XX 24-APR-2003.  
XX  
XX 11-OCT-2002; 2002WO-US032727.  
XX  
XX 15-OCT-2001; 2001US-00978825.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieue-Douglas J;  
XX  
XX WPI; 2003-381789/36.  
XX DR N-PSDB; ACF64489.  
XX  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
XX Example 1; SEQ ID NO 14713; 1481pp; English.  
XX  
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM5624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC polynucleotide of the invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising P. acnes polypeptides,  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 124 AA;  
SQ  
Query Match 87.2%; Score 34; DB 6; Length 124;  
Best Local Similarity 85.7%; Pred. No. 39;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LEHDGIN 7  
Db 34 LEHDGIS 40  
| | | | |  
RESULT 12  
AA39352  
ID AA39352 standard; peptide; 30 AA.  
XX  
XX AA39352;  
XX  
XX 13-DEC-1999 (first entry)  
XX  
XX Linker peptide consisting of six tetrapeptide caspase target sites.  
XX  
XX Fusion protein; repressor; linker; reporter; protease; cleavage site;  
KW

KW caspase; linker peptide; caspase; tetrapeptide; target site.  
 OS Synthetic.  
 XX WO9947640-A1.  
 PN 23-SEP-1999.  
 PD 19-MAR-1999; 99WO-US006070.  
 XX 20-MAR-1998; 98US-0078721P.  
 PF 17-MAR-1999; 99US-00270983.  
 PR (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 XX Hay BA, Hawkins CV;  
 PI WPI; 1999-571832/48.  
 XX A protease-dependent reporter fusion protein construct useful in assays  
 DR to identify proteases.  
 XX Example 1; Page 34; 64pp; English.  
 PS This sequence is a linker peptide consisting of six tetrapeptide caspase  
 XX target sites, each of which is followed by a glycine residue. The glycine  
 CC residues act as stabilizing residues in the N-end rule degradation  
 CC pathway in yeast. This linker sequence is an essential part of the fusion  
 CC protein of the invention. The invention relates to a protease-dependent  
 CC reporter-repressor fusion protein, where protease cleavage increases the  
 CC activity of the reporter. The fusion protein consists of a reporter  
 CC polypeptide linked to a linker polypeptide which is linked to a repressor  
 CC polypeptide. The linker polypeptide contains a protease cleavage site.  
 CC The repressor polypeptide represses the activity of the reporter  
 CC polypeptide and is operably linked to the linker polypeptide. Cleavage of  
 CC the linker polypeptide at the protease cleavage site increases the  
 CC activity of the reporter. The fusion proteins are useful in assays to  
 CC identify proteases that recognise specific protease cleavage sites or to  
 CC identify protease inhibitors or activators. The fusion proteins can be  
 CC used as reporters e.g. caspase activity molecules that function in living  
 CC cells  
 XX Sequence 30 AA;  
 SQ Query Match 79.5%; Score 31; DB 2; Length 30;  
 Best Local Similarity 83.3%; Pred. No. 35;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LEHGGI 6  
 DB :|||||  
 11 IEHDGI 16  
 RESULT 13  
 AAB27270  
 ID AAB27270 standard; peptide; 7 AA.  
 XX AC AAB27270;  
 XX 23-FEB-2001 (first entry)  
 DT Murine CRAM-1 tyrosine kinase phosphorylation signature #2.  
 DE Immunoglobulin superfamily; Ig Sf; C2 domain; vascular adhesion molecule;  
 XX inflammation; cancer; wound; angiogenesis; mouse;  
 KW confluency regulated adhesion molecule 1; CRAM-1; JAM-2.  
 XX Mus sp.  
 OS WO200053749-A2.  
 FN 14-SEP-2000.  
 XX The present invention describes fluorogenic compositions which can be

PF 13-MAR-2000; 2000WO-EP002219.  
 XX 11-MAR-1999; 99EP-00200746.  
 PR (RMFD-) RMF DICTAGENE SA.  
 XX Imhof BA, Aurrand-Lions M;  
 PI WPI; 2000-587436/55.  
 DR Isolated human Confluency Regulated Adhesion Molecule 1 or 2 (CRAM-1 or  
 XX CRAM-2) polypeptide, useful for treatment of tumors, inflammation  
 PT reactions and modulating vascular permeability.  
 XX Example; Page 23; 59pp; English.  
 PS The present sequence comprises one of the two tyrosine kinase  
 XX phosphorylation signatures found in the murine confluency regulated  
 CC adhesion molecule 1 (CRAM-1, also known as JAM-2). CRAM-1 is a member of  
 CC the immunoglobulin superfamily (Ig Sf). The CRAM-1 protein and coding  
 CC sequence can be used in the treatment of cancer, inflammation, to  
 CC modulate cell-cell interactions and angiogenesis, and in the modulation  
 CC of wound healing  
 XX Sequence 7 AA;  
 SQ Query Match 76.9%; Score 30; DB 3; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGIN 7  
 DB :|||||  
 1 KHDGVN 6  
 RESULT 14  
 AAG73290  
 ID AAG73290 standard; peptide; 7 AA.  
 XX AC AAG73290;  
 XX 14-AUG-2001 (first entry)  
 DT Protease indicator compound peptide #19.  
 DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
 XX viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
 KW haemophilia.  
 XX Synthetic.  
 OS WO200118238-A1.  
 PN 15-MAR-2001.  
 PD 11-SEP-2000; 2000WO-US024882.  
 PF 10-SEP-1999; 99US-00394019.  
 PR (ONCO-) ONCOIMMUNIN INC.  
 XX Komoriya A, Packard BS;  
 PI WPI; 2001-389573/41.  
 DR New fluorogenic compositions whose fluorescence level increases in the  
 XX presence of active proteases, useful for detecting and localizing  
 PT protease activity in biological samples, particularly in frozen tissue  
 PT samples.  
 XX Claim 1; Page 70; 86pp; English.  
 PS The present invention describes fluorogenic compositions which can be

CC used for the detection of protease activity. This can be useful as an  
 CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
 CC thrombosis and arthritis. The fluorogenic compositions comprise a  
 CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
 CC peptide is cleaved by a protease and the fluorophores can then be  
 CC detected. The present sequence is one of the peptides described in the  
 CC exemplification of the invention

XX Sequence 7 AA;

Query Match 76.9%; Score 30; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
 |||||  
 Db 3 HDGIN 7

#### RESULT 15

ABU60446  
 ID ABU60446 standard; peptide; 7 AA.

XX AC ABU60446;

XX DT 29-APR-2003 (first entry)

XX DE Protease binding peptide motif SEQ ID 170.

XX KW Protease; indicator; chromophore; H-dimer; fluorescence; absorbance;  
 KW nuclease; screening; fluorophore; substrate cleavage.

XX OS Synthetic.

XX PN WC200261038-A2.

XX PD 08-AUG-2002.

XX PF 21-DEC-2001; 2001WO-US049781.

XX PR 22-DEC-2000; 2000US-00747287.

XX PA (ONCO-) ONCOIMMUNIN INC.

XX PI Packard BS, Komoriya A;

XX PS WPI; 2002-698548/75.

XX PT Indicator composition comprising polypeptide or nucleic acid backbone  
 PT joining two same chromophores resulting in quenching of fluorescence  
 PT of/change in absorbance of chromophores, useful for detecting protease  
 PT activity.

XX PS Disclosure; Page 35; 97pp; English.

XX This invention describes a novel indicator composition (referred as homo-  
 CC doubly labeled compositions) comprising a polypeptide backbone or a  
 CC nucleic acid backbone joining two chromophores of the same species  
 CC whereby the chromophores form an H-dimer resulting in quenching of the  
 CC fluorescence of or a change in the absorbance of the chromophore, a  
 CC decrease in fluorescence or a change in absorbance indicates that the  
 CC first molecule and the second molecule are interacting. The indicator is  
 CC useful for detecting the activity of a protease, where an increase in  
 CC fluorescence or a change in absorbance indicates that the protease  
 CC cleaves the polypeptide backbone. The indicator is attached to a solid  
 CC support inside a mammalian, yeast or insect cell. The composition bears a  
 CC hydrophobic group such as Fmoc, 9-fluorenylmethyl group,<sup>1</sup>  
 CC fluorenylmethyl group, 9-fluorenylmethyl group, and 9-fluorenylmethyl-  
 CC carboxylic group, benzoyloxycarbonyl, Xanthyl (Xan), Trityl (Trt), 4-  
 CC methoxytrityl (Mtt), 4-methoxytrityl (Mmt), 4-methoxy-2,3, 6-trimethyl-  
 CC benzenesulphonyl (Mtr), Mesitylene-2-sulphonyl (Mts), 4,4'-  
 CC dimethoxybenzhydryl (Mbh), etc. The method described in the invention is  
 CC useful for detecting protease or nuclease activity (or the presence of

CC nucleic acid) in histological section, cells in culture, (e.g., seeded or  
 CC cultured adherent cells), a biological sample such as tissue, biopsy,  
 CC lymph, embryo, or whole animal, or cell suspension derived from a  
 CC biological sample such as tissue, blood, urine, saliva, lymph, or biopsy.  
 CC The indicator composition is also useful for screening a test agent for  
 CC the ability to modulate a protease (or a nuclease, lipase, etc.). The  
 CC indicator reagents allow rapid determination of protease activity in a  
 CC matter of minutes in a single-step procedure. The fluorescent indicators  
 CC both absorb and emit in the visible range (400-800 nm). These signals are  
 CC therefore not readily quenched by, nor is activation of the fluorophores,  
 CC that is, absorption of light, interfered with by background molecules;  
 CC therefore they are easily detected in biological samples. The fluorogenic  
 CC protease indicators utilize high efficiency fluorophores and are able to  
 CC achieve a high degree of quenching while providing a strong signal when  
 CC the quench is released by cleavage of the peptide substrate. The high  
 CC signal allows detection of very low levels of protease activity. Thus the  
 CC fluorogenic protease indicators are particularly well suited for in situ  
 CC detection of protease activity. ABU60357-ABU60477 represent peptides use  
 CC to illustrate the method described in the disclosure of the invention

XX Sequence 7 AA;

Query Match 76.9%; Score 30; DB 5; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
 |||||

Db 3 HDGIN 7

#### RESULT 16

AAG73328  
 ID AAG73328 standard; peptide; 18 AA.

XX AC AAG73328;

XX DT 14-AUG-2001 (first entry)

XX DE Protease indicator compound peptide #57.

XX KW Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
 KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
 KW haemophilia.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 4

FT /label= OTHER

FT /note= "designated J in the specification"

FT Modified-site 14

FT /label= OTHER

FT /note= "designated J in the specification"

XX WO200118238-A1.

XX PD 15-MAR-2001.

XX PF 11-SEP-2000; 2000WO-US024882.

XX PR 10-SEP-1999; 99US-00394019.

XX PA (ONCO-) ONCOIMMUNIN INC.

XX PI Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

XX New fluorogenic compositions whose fluorescence level increases in the  
 PT presence of active proteases, useful for detecting and localizing  
 PT protease activity in biological samples, particularly in frozen tissue  
 PT samples.

XX PS Claim 4; Page 71; 86pp; English.

XX CC The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention

XX SQ Sequence 18 AA;

Query Match 76.9%; Score 30; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HDGIN 7  
Db 8 HDGIN 12  
|||||

RESULT 17

AAAG73234  
ID AAG73234 standard; peptide; 18 AA.  
XX AC AAG73234;  
XX AC AAG73234;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease binding site #168.  
XX KW Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
XX KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
XX KW haemophilia.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
FT Modified-site 1 /label= OTHER  
FT FT /note= "modified by fluorophore"  
XX WO200118238-A1.  
XX PD 15-MAR-2001.  
XX PF 11-SEP-2000; 2000WO-US024882.  
XX PR 10-SEP-1999; 99US-00394019.  
XX FA (ONCO-) ONCOMUNIN INC.  
XX FI Komoriya A, Packard BS;  
XX DR WPI; 2001-389573/41.  
XX PT New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples.  
XX PS Disclosure; Page 29; 86pp; English.

XX CC The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention

XX SQ Sequence 18 AA;

Query Match 76.9%; Score 30; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HDGIN 7  
Db 8 HDGIN 12  
|||||

RESULT 18

AAAY15690  
ID AAY15690 standard; peptide; 5 AA.  
XX AC AAY15690;  
XX DT 27-JUL-1999 (first entry)  
XX DE Peptide used to make fluorescent reporter molecules.  
XX KW Fluorogenic; fluorescent reporter molecule; enzyme substrate; apoptosis;  
XX KW protease; peptidase; apoptosis cascade; cancer; chemotherapeutic agent;  
XX KW cell death; viral protease activity.  
XX OS Synthetic.  
XX PN WO9918856-A1.  
XX PD 22-APR-1999.  
XX PF 09-OCT-1998; 98WO-US021231.  
XX PR 10-OCT-1997; 97US-0061582P.  
XX PR 03-MAR-1998; 98US-00033661.  
XX PA (CYTO-) CYTOVIA INC.  
XX PI Weber E, Cai SX, Keana JFW, Drewe JA, Zhang H;  
XX WPI; 1999-312448/26.  
XX PT New fluorogenic or fluorescent reporter molecules.  
XX PS Disclosure; Page 179; 202pp; English.

XX CC AAY15618-Y15759 represent peptides used to make the fluorogenic or fluorescent reporter molecules of the invention. These molecules contain a peptide moiety (e.g. present sequence) which acts as a substrate for enzymes involved in apoptosis or protease or peptidase enzymes. The compounds can be used as fluorogenic or fluorescent substrates for enzymes. Depending on the peptide moiety used, the fluorescent molecules can be used for detecting or measuring the activity of an enzyme involved in the apoptosis cascade in cells, to determine whether a test compound has an effect on an enzyme involved in the apoptosis cascade in cells; for determining the sensitivity of an animal with cancer to treatment with chemotherapeutic agents or determining whether a test substance inhibits, prevents, causes or enhances cell death of test cells; for detecting or measuring the activity of a viral protease in cells; for determining whether a test compound has an effect on the activity of a viral protease in cells; and for measuring the activity or determining whether a test substance has an effect on the activity of a protease or peptidase in cells

XX SQ Sequence 5 AA;

Query Match 74.4%; Score 29; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDG 5  
|||||

Db 1 LEHDG 5

RESULT 19

AAG73287 standard; peptide; 7 AA.

AC AAG73287;

DT 14-AUG-2001 (first entry)

XX Protease indicator compound peptide #16.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;

KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;

KW haemophilia.

OS Synthetic.

XX WO200118238-A1.

PN 15-MAR-2001.

PD 11-SEP-2000; 2000WO-US024882.

PF 10-SEP-1999; 99US-00394019.

PR (ONCO-) ONCOIMMUNIN INC.

PA Komoriya A, Packard BS;

PI WPI; 2001-389573/41.

DR New fluorogenic compositions whose fluorescence level increases in the

XX presence of active proteases, useful for detecting and localizing

PT protease activity in biological samples, particularly in frozen tissue

PT samples.

XX Claim 1; Page 70; 86pp; English.

PS The present invention describes fluorogenic compositions which can be

XX used for the detection of protease activity. This can be useful as an

CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,

CC thrombosis and arthritis. The fluorogenic compositions comprise a

CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The

CC peptide is cleaved by a protease and the fluorophores can then be

CC detected. The present sequence is one of the peptides described in the

CC exemplification of the invention

XX Sequence 7 AA;

Query Match 74.4%; Score 29; DB 4; Length 7;

Best Local Similarity 85.7%; Pred. No. 1.4e+06;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDGIN 7

Db 1 LETDGIN 7

RESULT 20

ABU60443

ID ABU60443 standard; peptide; 7 AA.

XX ABU60443;

AC 29-APR-2003 (first entry)

DT Protease binding peptide motif SEQ ID 167.

XX Protease; indicator; chromophore; H-dimer; fluorescence; absorbance;

KW nuclease; screening; fluorophore; substrate cleavage.

XX

OS Synthetic.

XX WO200261038-A2.

XX 08-AUG-2002.

PD 21-DEC-2001; 2001WO-US049781.

PF 22-DEC-2000; 2000US-00747287.

PR (ONCO-) ONCOIMMUNIN INC.

PA Packard BS, Komoriya A;

PI WPI; 2002-698548/75.

DR Indicator composition comprising polypeptide or nucleic acid backbone

XX joining two same chromophores resulting in quenching of fluorescence

PT of/change in absorbance of chromophores, useful for detecting protease

PT activity.

XX Disclosure; Page 35; 97pp; English.

PS This invention describes a novel indicator composition (referred as homo-

XX doubly labeled compositions) comprising a polypeptide backbone or a

CC nucleic acid backbone joining two chromophores of the same species

CC whereby the chromophores form an H-dimer resulting in quenching of the

CC fluorescence of or a change in the absorbance of the chromophore, a

CC decrease in fluorescence or a change in absorbance indicates that the

CC first molecule and the second molecule are interacting. The indicator is

CC useful for detecting the activity of a protease, where an increase in

CC fluorescence or a change in absorbance indicates that the protease

CC cleaves the polypeptide backbone. The indicator is attached to a solid

CC support inside a mammalian, yeast or insect cell. The composition bears a

CC hydrophobic group such as Fmoc, 9-fluorenesulfonyl group, 1-fluorenone-1-

CC fluorinecarboxylic group, 9-fluoreneacetyl group, and 4-fluorenone-1-

CC carboxylic group, benzyloxycarbonyl, Xanthyl (Xan), Trityl (Trt), 4-

CC methyltrityl (Mtt), 4-methoxytrityl (Mmt), 4-methoxy-2,3,6-trimethyl-

CC benzene-sulphonyl (Mtr), Mesitylene-2-sulphonyl (Mts), 4,4'-

CC dimethoxybenzhydryl (Mbh), etc. The method described in the invention is

CC useful for detecting protease or nuclease activity (or the presence of

CC nucleic acid) in histological section, cells in culture, (e.g., seeded or

CC cultured adherent cells), a biological sample such as tissue, biopsy,

CC lymph, embryo, or whole animal, or cell suspension derived from a

CC biological sample such as tissue, blood, urine, saliva, lymph, or biopsy.

CC The indicator composition is also useful for screening a test agent for

CC the ability to modulate a protease (or a nuclease, lipase, etc.). The

CC indicator reagents allow rapid determination of protease activity in a

CC matter of minutes in a single-step procedure. The fluorescent indicators

CC both absorb and emit in the visible range (400-800 nm). These signals are

CC therefore not readily quenched by, nor is activation of the fluorophores,

CC that is, absorption of light, interfered with by background molecules;

CC therefore they are easily detected in biological samples. The fluorogenic

CC protease indicators utilise high efficiency fluorophores and are able to

CC achieve a high degree of quenching while providing a strong signal when

CC the quench is released by cleavage of the peptide substrate. The high

CC signal allows detection of very low levels of protease activity. Thus the

CC fluorogenic protease indicators are particularly well suited for in situ

CC detection of protease activity. ABU60357-ABU60477 represent peptides use

CC to illustrate the method described in the disclosure of the invention

XX Sequence 7 AA;

Query Match 74.4%; Score 29; DB 5; Length 7;

Best Local Similarity 85.7%; Pred. No. 1.4e+06;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDGIN 7

Db 1 LETDGIN 7

RESULT 21

AAG73231  
ID AAG73231 standard; peptide; 18 AA.  
AC AAG73231;  
XX  
XX 14-AUG-2001 (first entry)  
XX  
XX Protease binding site #165.  
XX  
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
KW haemophilia.  
XX  
XX Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FH Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"  
FT  
FN WO200118238-A1.  
XX  
XX 15-MAR-2001.  
PD  
XX  
XX 11-SEP-2000; 2000WO-US024882.  
FF  
XX  
XX 10-SEP-1999; 99US-00394019.  
PR  
XX (ONCO-) ONCOIMMUNIN INC.  
PA  
XX Komoriya A, Packard BS;  
PI  
XX WPI; 2001-389573/41.  
DR  
XX New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples.  
PT  
XX Disclosure; Page 29; 86pp; English.  
PS  
XX The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention  
XX  
SQ Sequence 18 AA;  
Query Match 74.4%; Score 29; DB 4; Length 18;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LEHDGIN 7  
DB 6 LETDGIN 12  
RESULT 22  
AAG73325  
ID AAG73325 standard; peptide; 18 AA.  
XX  
XX AAG73325;  
AC  
XX  
XX 14-AUG-2001 (first entry)  
DT  
XX Protease indicator compound peptide #54.  
DE  
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis; thrombosis;  
KW

haemophilia.  
KW  
XX Synthetic.  
OS  
XX Key Location/Qualifiers  
FH Modified-site 4  
FT /label= OTHER  
FT /note= "designated J in the specification"  
FT  
FT Modified-site 14  
FT /label= OTHER  
FT /note= "designated J in the specification"  
XX  
XX WO200118238-A1.  
PN  
XX  
XX 15-MAR-2001.  
PD  
XX  
XX 11-SEP-2000; 2000WO-US024882.  
PF  
XX  
XX 10-SEP-1999; 99US-00394019.  
PR  
XX (ONCO-) ONCOIMMUNIN INC.  
XX  
XX Komoriya A, Packard BS;  
PI  
XX WPI; 2001-389573/41.  
DR  
XX New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples.  
PT  
XX Claim 4; Page 71; 86pp; English.  
PS  
XX The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention  
XX  
SQ Sequence 18 AA;  
Query Match 74.4%; Score 29; DB 4; Length 18;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LEHDGIN 7  
DB 6 LETDGIN 12  
RESULT 23  
ABB9029  
ID ABB9029 standard; peptide; 20 AA.  
XX  
XX ABB9029;  
AC  
XX  
XX 20-JUN-2002 (first entry)  
DT  
XX  
XX Babesia microti antigenic epitope fusion protein peptide BMN N term 4.  
DE  
XX Protozoacide; vaccine; antigen; antigenic epitope; infection.  
KW  
XX Babesia microti.  
OS  
XX WO200185947-A2.  
PN  
XX  
XX 15-NOV-2001.  
PD  
XX  
XX 09-MAY-2001; 2001WO-US015192.  
PF  
XX

```

PA (HOUG/) HOUGHTON R L.
PA (SLEA/) SLEATH P R.
PA (MCNE/) MCNEILL P D.
PA (HOME/) HOMER M J.
PA (SECR/) SECRIST H.
XX
XX Read SG, Lodes MJ, Houghton RL, Sleath PR, Mcneill PD, Homer MJ;
PI Secrist H;
XX
XX WPI; 2003-801279/75.
XX
XX New isolated polynucleotide for stimulating an immune response in a
XX patient and/or for treating or detecting a Babesia microti infection.
XX
XX Claim 10; SEQ ID NO 215; 167pp; English.
XX
XX The invention relates to Babesia microti polynucleotides and the
XX polypeptides they encode. The polypeptides, antibodies specific for the
XX polypeptides and fusion proteins comprising the polypeptides are used
XX with immunostimulants for stimulating an immune response in a patient and
XX for treating a Babesia microti infection in a patient. Oligonucleotides
XX are used to diagnose B. microti infection in a patient by contacting a
XX biological sample from the patient with the oligonucleotide, detecting an
XX amount of binding of a polynucleotide that hybridises to the
XX oligonucleotide and comparing the amount of the polynucleotide that
XX hybridises to the oligonucleotide to a cut-off value. This sequence
XX represents a Babesia microti antigen (BMNI)-related peptide of the
XX invention.
XX
XX Sequence 20 AA;
XX
XX Query Match 74.4%; Score 29; DB 7; Length 20;
XX Best Local Similarity 83.3%; Pred No. 57;
XX Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 2 EHDGIN 7
XX |||||
XX Db 6 EHDGIN 11
XX
XX
XX RESULT 25
XX AAG07964
XX ID AAG07964 standard; protein; 38 AA.
XX
XX AC AAG07964;
XX
XX XX
XX DT 17-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 5316.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126264P.
XX PR 29-MAR-1999; 99US-0126785P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129845P.
XX PR 19-APR-1999; 99US-0130077P.

```



PR	20-JUL-1999	99US-01443552Z
PR	20-JUL-1999	99US-01446322P
PR	20-JUL-1999	99US-01448844P
PR	21-JUL-1999	99US-01448144P
PR	21-JUL-1999	99US-01450866P
PR	21-JUL-1999	99US-01450888P
PR	22-JUL-1999	99US-01450858P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145089P
PR	22-JUL-1999	99US-01451922P
PR	23-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-01452148P
PR	23-JUL-1999	99US-01452242P
PR	23-JUL-1999	99US-0145276P
PR	23-JUL-1999	99US-01459133P
PR	24-JUL-1999	99US-0145918P
PR	27-JUL-1999	99US-01459198P
PR	28-JUL-1999	99US-01459511P
PR	28-JUL-1999	99US-01459514P
PR	02-AUG-1999	99US-01463882P
PR	03-AUG-1999	99US-01463899P
PR	03-AUG-1999	99US-01470388P
PR	04-AUG-1999	99US-01472042P
PR	05-AUG-1999	99US-01471922P
PR	06-AUG-1999	99US-0147303P
PR	08-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-01479355P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-01483419P
PR	12-AUG-1999	99US-01483411P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-01493688P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	23-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149902P
PR	23-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	30-AUG-1999	99US-01510802P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151438P
PR	01-SEP-1999	99US-0151930P
PR	07-SEP-1999	99US-0152363P
PR	10-SEP-1999	99US-0153070P
PR	13-SEP-1999	99US-0153758P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0154779P
PR	22-SEP-1999	99US-0155139P
PR	23-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155659P
PR	28-SEP-1999	99US-0156458P
PR	29-SEP-1999	99US-0156596P
PR	04-OCT-1999	99US-0157117P
PR	05-OCT-1999	99US-0157753P
PR	06-OCT-1999	99US-0157865P
PR	07-OCT-1999	99US-0158029P
PR	08-OCT-1999	99US-0158232P
PR	12-OCT-1999	99US-0158369P
PR	13-OCT-1999	99US-0158293P
PR	13-OCT-1999	99US-0158944P
PR	14-OCT-1999	99US-0159295P
PR	14-OCT-1999	99US-0159329P
PR	14-OCT-1999	99US-0159330P

PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159337P.  
PR 14-OCT-1999; 99US-0159338P.  
PR 18-OCT-1999; 99US-0159384P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 74.4%; Score 29; DB 3; Length 38;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDEGI 6  
Db 9 LEHDEGI 14

RESULT 26  
AAV55707  
ID AAV55707 standard; protein; 39 AA.  
XX AC AAV55707;  
XX DT 07-FEB-2000 (first entry)  
XX DE M. leprae ycfB peptide motif 4.  
XX KW ycfB polypeptide; gram negative bacteria; gram positive bacteria;  
XX KW antibacterial; bacterial infection; bacterial viability; antibiotic.  
XX OS Mycobacterium leprae.  
XX PN WO9954462-A2.  
XX PD 28-OCT-1999.  
XX PF 20-APR-1999; 99WO-EP002638.  
XX PR 22-APR-1998; 98GB-00008350.  
XX PA (GLAX) GLAXO GROUP LTD.  
XX PI Arigoni F, Edgerton MD, Loferer H, Peitsch MC;  
XX DR WPI; 2000-013249/01.  
XX PT Novel bacterial polypeptides used to identify broad spectrum antibiotics.  
XX PS Claim 3; Fig 2D; 56pp; English.  
CC The invention discloses a novel family of polypeptides, designated the ycfB family, required for the growth of both gram negative and gram positive bacteria. The novel polypeptide of ycfB family is defined by:  
CC (a) a HSP score of greater than or equal to 100 when compared with one of the amino acid sequences of the ycfB family members given in the specification, when the BLAST algorithm is used with a BLOSUM62 scoring matrix; (b) containing a set of amino acid sequences which are positively

CC identified when position dependent scoring matrices according to tables 1-4 (given in the specification) are used to yield a p-value of less than 110-60; or (c) comprising any one of the sequences shown in AAY55710-711. The ycfB polypeptides and polynucleotides can be used in method to identify antagonists and antibacterial compounds. These antagonists and compounds can be used to treat bacterial infections. The polypeptides of the invention are essential proteins for bacterial viability, and represent new targets for antibiotics. Sequences AAY5647-709 represent different motifs of the ycfB family polypeptides of various bacterial species

XX Sequence 39 AA;  
SQ

Query Match 74.4%; Score 29; DB 3; Length 39;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
Db 13 EHDGVH 18

RESULT 27  
AAW60953  
ID AAW60953 standard; protein; 40 AA.  
XX AC AAW60953;  
XX DT 13-OCT-1998 (first entry)  
XX DE Streptococcus pneumoniae encoded polypeptide.  
XX KW coding region; ORF; open reading frame; antibacterial; infection; prevention; meningitis.  
XX KW Streptococcus pneumoniae.  
XX OS Streptococcus pneumoniae.  
XX PN WO9819689-A1.  
XX PD 14-MAY-1998.  
XX PF 27-OCT-1997; 97WO-US019226.  
XX PR 01-NOV-1996; 96US-0029930P.  
XX PA (SMIK) SMITHKLINE BEECHAM CORP.  
XX PA (SMIK) SMITHKLINE BEECHAM PLC.  
XX PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
XX PI Reid RH, Zarfos PN;  
XX DR WPI; 1998-286586/25.  
XX DR N-PSDB; AAV37353.  
XX PT New isolated nucleic acids from Streptococcus pneumoniae - useful, e.g. for identifying anti-bacterial(s) for treatment and prevention of meningitis.  
XX PS Claim 11; Page 54; 130pp; English.  
CC The sequence is that of the polypeptide encoded by a region isolated from S. pneumoniae which shows homology to glutamate racemase. The protein, or agonists of it, may be useful as an antibacterial for treatment or prevention of infection, specifically caused by S. pneumoniae (particularly meningitis) but possibly also Helicobacter pylori (ulcers and gastric cancer). It may be of particular use before insertion of an in-dwelling device or any other invasive procedure. The protein, or nucleic acid encoding it, can also be used in vaccines to induce a cellular and/or humoral immune response, or to screen for other antibacterials. The DNA may also contain flanking sequences that are potential sources of control elements for bacterial gene expression. Detecting a sequence encoding the protein can be used diagnostically, e.g. to detect a mutation for serotyping or classifying infectious agents



PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	15-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149930P.
PR	23-AUG-1999;	99US-0150566P.
PR	25-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155559P.
PR	28-SEP-1999;	99US-0156458P.
PR	28-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	08-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158322P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161406P.
PR	25-OCT-1999;	99US-0161405P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161933P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 74.4%; Score 29; DB 3; Length 52;		
Best Local Similarity 83.3%; Pred.No. 1.6e+02;		
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;		

QY	1 LEHDCI 6
Db	23 LEHEGI 28
RESULT 29	
AAG49408	
ID	AAG49408 standard; protein; 53 AA.
XX	AAG49408;
XX	18-OCT-2000 (first entry)
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 62503.
XX	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	Arabidopsis thaliana.
XX	EP1033405-A2.
XX	06-SEP-2000.
XX	25-FEB-2000; 2000EP-00301439.
PR	25-FEB-1999; 99US-0121825P.
PR	05-MAR-1999; 99US-0123180P.
PR	09-MAR-1999; 99US-0123548P.
PR	23-MAR-1999; 99US-0125788P.
PR	25-MAR-1999; 99US-0126264P.
PR	29-MAR-1999; 99US-0126785P.
PR	01-APR-1999; 99US-0127462P.
PR	06-APR-1999; 99US-0128234P.
PR	08-APR-1999; 99US-0128714P.
PR	16-APR-1999; 99US-0129845P.
PR	19-APR-1999; 99US-0130077P.
PR	21-APR-1999; 99US-0130449P.
PR	23-APR-1999; 99US-0130510P.
PR	23-APR-1999; 99US-0130891P.
PR	28-APR-1999; 99US-0131449P.
PR	30-APR-1999; 99US-0132048P.
PR	30-APR-1999; 99US-0132407P.
PR	04-MAY-1999; 99US-0132484P.
PR	05-MAY-1999; 99US-0132485P.
PR	06-MAY-1999; 99US-0132486P.
PR	06-MAY-1999; 99US-0132487P.
PR	07-MAY-1999; 99US-0132863P.
PR	11-MAY-1999; 99US-0134256P.
PR	14-MAY-1999; 99US-0134218P.
PR	14-MAY-1999; 99US-0134219P.
PR	14-MAY-1999; 99US-0134221P.
PR	14-MAY-1999; 99US-0134370P.
PR	18-MAY-1999; 99US-0134768P.
PR	19-MAY-1999; 99US-0134941P.
PR	20-MAY-1999; 99US-0135124P.
PR	21-MAY-1999; 99US-0135353P.
PR	24-MAY-1999; 99US-0135629P.
PR	25-MAY-1999; 99US-0136021P.
PR	27-MAY-1999; 99US-0136392P.
PR	28-MAY-1999; 99US-0136782P.
PR	01-JUN-1999; 99US-0137222P.
PR	03-JUN-1999; 99US-0137528P.
PR	04-JUN-1999; 99US-0137502P.
PR	07-JUN-1999; 99US-0137724P.
PR	08-JUN-1999; 99US-0138094P.
PR	10-JUN-1999; 99US-0138540P.
PR	10-JUN-1999; 99US-0138847P.
PR	14-JUN-1999; 99US-0139113P.
PR	16-JUN-1999; 99US-0139452P.
PR	16-JUN-1999; 99US-0139453P.
PR	17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 23-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140696P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143543P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147193P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.

PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 22-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 25-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 08-OCT-1999; 99US-0157865P.  
PR 08-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158432P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 23-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 74.4%; Score 29; DB 3; Length 53;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDCI 6  
|||:  
Db 9 LEHEGI 14

RESULT 30  
AAR37963

ID XX AAR37963 standard; protein; 64 AA.  
AC XX AAR37963;  
XX XX  
DT 25-MAR-2003 (revised)  
DT 07-OCT-1993 (first entry)  
XX XX  
DE KGFR Ig-like domain 3 V-region inserted into FGFR2 framework.  
XX XX  
KW Polymerase chain reaction; immunoglobulin-like domain; tumour;  
KW ligand specificity region; LSK; fibroblast growth factor receptor;  
KW variable region; keratinocyte growth factor receptor; chimera.  
XX XX  
OS Mus musculus.  
XX XX  
FH Key Location/Qualifiers  
FT Region 1..52  
FT Region /note= "mouse KGFR Ig-like domain 3 V-region"  
FT Region 53..64  
FT Region /note= "FGFR2 framework"  
XX XX  
FN EP545343-A1.  
XX XX  
PD 09-JUN-1993.  
XX XX  
PF 30-NOV-1992; 92EP-00120425.  
XX XX  
PR 02-DEC-1991; 91IL-00100219.  
XX XX  
PA (YEDA ) YEDA RES & DEV CO LTD.  
XX XX  
PI Yayon A, Givol D;  
XX XX  
DR WPI; 1993-184106/23.  
DR N-PSDB; AAR37963.  
XX XX  
PT Ligand-binding domains of fibroblast growth factor receptor - used in  
PT chimeric proteins, inhibit proliferation of tumour cells.  
XX XX  
PS Example 1; Fig 3; 20pp; English.  
XX XX  
CC To prepare chimeric KGFR/FGFR2, a cDNA corresponding to the KGFR-v region  
CC of Ig-like domain D3 was cloned from mouse skin RNA using the PCR primer  
CC AAQ43158. This primer overlaps the constant region and the KGFR variable  
CC region of D3. Sequence analysis of the PCR product and the chimeric  
CC construct confirmed their identity to the KGFR sequence in that region  
CC and the in-frame insertion of the variable segment (see AAQ43163 and  
CC AAR37963). (Updated on 25-MAR-2003 to correct PN field.)  
XX XX  
SQ Sequence 64 AA;  
XX XX  
Query Match 74.4%; Score 29; DB 2; Length 64;  
Best Local Similarity 71.4%; Pred.No.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LEHDSIN 7  
Db 1 LKHSGIN 7  
XX XX  
RESULT 31  
AAM84389  
ID AAM84389 standard; protein; 66 AA.  
XX XX  
AC AAM84389;  
XX XX  
DT 07-NOV-2001 (first entry)  
DE Human immune/haematopoietic antigen SEQ ID NO:11982.  
XX XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
XX XX

OS Homo sapiens.  
XX XX  
PN WO200157182-A2.  
XX XX  
PD 09-AUG-2001.  
XX XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236337P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249254P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249257P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0255967P.  
(HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-483426/52.  
DR N-FSDB; AAK57170.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX Claim 11; SEQ ID NO 11982; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytosstatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX SQ Sequence 66 AA;  
Query Match 74.4%; Score 29; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Mismatches 0; Gaps 0;  
Matches 5; Conservative 0; Indels 0;  
QY 1 LEHDG 5  
Db |||||  
56 LEHDG 60  
RESULT 32  
AAG49407  
ID AAG49407 standard; protein; 67 AA.  
XX AAG49407;  
AC AAG49407;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 62502.  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62502.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX RP1033405-A2.  
PN 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
PP 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0125264P.  
PR 29-MAR-1999; 99US-0125785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.





PR	05-MAY-1999;	90US-0132485P;
PR	06-MAY-1999;	90US-0132486P;
PR	06-MAY-1999;	90US-0132487P;
PR	07-MAY-1999;	90US-0132863P;
PR	11-MAY-1999;	90US-0134256P;
PR	14-MAY-1999;	90US-0134219P;
PR	14-MAY-1999;	90US-0134219P;
PR	14-MAY-1999;	90US-0134221P;
PR	14-MAY-1999;	90US-0134370P;
PR	18-MAY-1999;	90US-0134768P;
PR	19-MAY-1999;	90US-0134941P;
PR	20-MAY-1999;	90US-0135124P;
PR	21-MAY-1999;	90US-0135353P;
PR	24-MAY-1999;	90US-0135629P;
PR	25-MAY-1999;	90US-0136021P;
PR	27-MAY-1999;	90US-0136392P;
PR	28-MAY-1999;	90US-0136782P;
PR	01-JUN-1999;	90US-0137222P;
PR	03-JUN-1999;	90US-0137528P;
PR	04-JUN-1999;	90US-0137502P;
PR	07-JUN-1999;	90US-0137724P;
PR	08-JUN-1999;	90US-0138094P;
PR	10-JUN-1999;	90US-0138340P;
PR	10-JUN-1999;	90US-0138847P;
PR	14-JUN-1999;	90US-0139119P;
PR	16-JUN-1999;	90US-0139452P;
PR	16-JUN-1999;	90US-0139453P;
PR	17-JUN-1999;	90US-0139459P;
PR	18-JUN-1999;	90US-0139460P;
PR	18-JUN-1999;	90US-0139461P;
PR	18-JUN-1999;	90US-0139462P;
PR	18-JUN-1999;	90US-0139463P;
PR	18-JUN-1999;	90US-0139750P;
PR	18-JUN-1999;	90US-0139763P;
PR	21-JUN-1999;	90US-0139817P;
PR	22-JUN-1999;	90US-0139899P;
PR	23-JUN-1999;	90US-0140353P;
PR	23-JUN-1999;	90US-0140354P;
PR	24-JUN-1999;	90US-0140695P;
PR	24-JUN-1999;	90US-0140695P;
PR	28-JUN-1999;	90US-0140823P;
PR	29-JUN-1999;	90US-0140991P;
PR	08-JUL-1999;	90US-0142803P;
PR	09-JUL-1999;	90US-0142920P;
PR	12-JUL-1999;	90US-0142977P;
PR	13-JUL-1999;	90US-0143542P;
PR	14-JUL-1999;	90US-0143624P;
PR	02-JUL-1999;	90US-0144005P;
PR	06-JUL-1999;	90US-0144005P;
PR	16-JUL-1999;	90US-0144085P;
PR	16-JUL-1999;	90US-0144086P;
PR	19-JUL-1999;	90US-0144325P;
PR	19-JUL-1999;	90US-0144331P;
PR	19-JUL-1999;	90US-0144332P;
PR	20-JUL-1999;	90US-0144332P;
PR	21-JUL-1999;	90US-0144804P;
PR	21-JUL-1999;	90US-0144814P;
PR	21-JUL-1999;	90US-0145086P;
PR	22-JUL-1999;	90US-0145088P;
PR	22-JUL-1999;	90US-0145085P;

PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145122P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-01451145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145949P.	PR	26-OCT-1999;	99US-0161359P.
PR	28-JUL-1999;	99US-0145961P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146388P.	PR	28-OCT-1999;	99US-0161920P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161992P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147204P.	PR	29-OCT-1999;	99US-0162142P.
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147132P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	23-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150586P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156456P.			
PR	29-SEP-1999;	99US-0156596P.			
PR	04-OCT-1999;	99US-0157117P.			
PR	05-OCT-1999;	99US-0157533P.			
PR	06-OCT-1999;	99US-0157855P.			
PR	07-OCT-1999;	99US-0158029P.			
PR	08-OCT-1999;	99US-0158232P.			
PR	12-OCT-1999;	99US-0158369P.			
PR	13-OCT-1999;	99US-0159293P.			
PR	13-OCT-1999;	99US-0159294P.			
PR	13-OCT-1999;	99US-0159295P.			
PR	14-OCT-1999;	99US-0159329P.			
PR	14-OCT-1999;	99US-0159330P.			
PR	14-OCT-1999;	99US-0159331P.			
PR	14-OCT-1999;	99US-0159637P.			
PR	14-OCT-1999;	99US-0159638P.			
PR	18-OCT-1999;	99US-0159584P.			
PR	21-OCT-1999;	99US-0160741P.			
PR	21-OCT-1999;	99US-0160767P.			
PR	21-OCT-1999;	99US-0160768P.			

Query Match

Best Local Similarity

Matches

5; Conservative

1; Mismatches

0; Indels

0; Gaps

0;

QY

1 LEHDGI 6

DB

3 LEHEGI 8

RESULT 34

ABP80959

ID ABP80959 standard; protein; 82 AA.

XX

AC ABP80959;

XX

DT 07-MAR-2003 (first entry)

XX

DE N. gonorrhoeae amino acid sequence SEQ ID 8448.

XX

KW Antibacterial; infection; vaccine; gene therapy.

XX

OS Neisseria gonorrhoeae.

XX

PN WO200279243-A2.

XX

PD 10-OCT-2002.

XX

PF 12-FEB-2002; 2002WO-IB002069.

XX

PR 12-FEB-2001; 2001GB-00003424.

XX

PA (CHIR-) CHIRON SPA.

XX

PI Fontana MR, Pizza M, Masignani V, Monaci E;

XX

PI WPI; 2003-058415/05.

XX

DR N-PSDE; AB241929.

XX

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection.

PS Disclosure; Page 803; 815pp; English.

XX

CC The present invention relates to proteins from Neisseria gonorrhoeae.

CC Also disclosed are the nucleic acid molecules encoding the proteins and antibodies that specifically bind to the proteins. The composition comprising the protein, nucleic acid or antibody is useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid molecules of the invention

XX

Sequence 82 AA;

Query Match

Best Local Similarity

74.4%; Score 29; DB 6; Length 82;

66.7%; Pred. NO. 2.6e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 BHDGIN 7  
DB 64 BHGVN 69

RESULT 35  
ADC95330  
ID ADC95330 standard; protein; 83 AA.

XX AC ADC95330;  
XX 01-JAN-2004 (first entry)  
XX E. faecium protein sequence SEQ ID 4957.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
KW abdominal-pelvic infection.

XX Enterococcus faecium.

XX US6583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

XX 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2003-799836/75.

XX N-PSDB; ADC91676.

XX New isolated nucleic acid derived from Enterococcus faecium encoding an  
PT Enterococcus faecium polypeptide useful for detection, prevention and  
PT treatment of a pathological condition resulting from a bacterial  
PT infection.

XX Example 1; SEQ ID NO 4957; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridising to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids are useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract  
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of Candida albicans -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one if the disclosed E. faecium proteins.

XX Sequence 83 AA;

Query Match 74.4%; Score 25; DB 7; Length 83;  
Best Local Similarity 83.3%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGI 6  
DB 57 LKHDGI 52

RESULT 36  
ABP64017  
ID ABP64017 standard; protein; 84 AA.

XX AC ABP64017;  
XX 04-NOV-2002 (first entry)  
XX Human ORF387.

XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;  
KW Antinflammatory; gene therapy; human; ORFX; atherogenic; platelet;  
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;  
KW cancer; cardiovascular disease; allergy; autoimmune disease;  
KW wound healing; blood coagulation disorder; inflammatory disorder.

XX Homo sapiens.

XX US2002082206-A1.

XX 27-JUN-2002.

XX 30-MAY-2001; 2001US-00857550.

XX 30-MAY-2000; 2000US-0208427P.

XX (LEAC/) LEACH M D.

XX (MEHR/) MEHRABAN F.

XX (CONL/) CONLEY P B.

XX (TOPP/) TOPPER J N.

XX (LAWD/) LAW D.

XX Leach MD, Mehraban F, Conley PB, Topper JN, Law D;

XX WPI; 2002-626554/67.

XX N-PSDB; ABQ98580.

XX New polypeptide designated ORFX are present in human atherogenic cells  
PT and are useful to prevent and treat ORFX-associated disorders including  
PT cancer, allergy, wound healing or autoimmune, cardiovascular or  
PT inflammatory disease.

XX Claim 10; SEQ ID NO 774; 78pp; English.

XX The present invention relates to novel human ORFX polypeptides and their  
CC coding sequences (ABP6361-ABP6461 and ABQ98194-ABQ99267). The sequences  
CC were discovered in human atherogenic cells, in particular in platelets  
CC and human umbilical vein endothelial cells (HUVEC) and are expressed in  
CC many other tissues as well. Atherogenic cells are cells which have the  
CC potential to develop atherosclerotic plaques. The ORFX polypeptides and  
CC nucleic acids are useful for treating or preventing a pathological  
CC condition associated with an ORFX-associated disorder, e.g. cancer, blood  
CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood  
CC coagulation disorders or inflammatory disorders. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from the USPTO web site at  
CC seqdata.uspto.gov/sequence.html?DocID=20020082206

XX Sequence 84 AA;

Query Match 74.4%; Score 29; DB 5; Length 84;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
DB 29 LEHDG 33

RESULT 37  
AAG34646

ID	AG34646 standard; protein; 87 AA.	PR	18-JUN-1999;	99US-0139461P
XX		PR	18-JUN-1999;	99US-0139462P
AC	AG34646;	PR	18-JUN-1999;	99US-0139463P
XX		PR	18-JUN-1999;	99US-0139750P
DT		PR	18-JUN-1999;	99US-0139763P
XX		PR	21-JUN-1999;	99US-0139817P
XX	18-OCT-2000 (first entry)	PR	22-JUN-1999;	99US-0139899P
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 42191.	PR	23-JUN-1999;	99US-0140353P
XX		PR	23-JUN-1999;	99US-0140354P
KW	Protein identification; signal transduction pathway; metabolic pathway;	PR	24-JUN-1999;	99US-0140695P
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	28-JUN-1999;	99US-0140823P
KW	termination sequence.	PR	29-JUN-1999;	99US-0140991P
XX		PR	30-JUN-1999;	99US-0141287P
OS	Arabidopsis thaliana.	PR	01-JUL-1999;	99US-0141842P
XX		PR	01-JUL-1999;	99US-0142154P
PN	EP1033405-A2.	PR	02-JUL-1999;	99US-0142055P
XX		PR	06-JUL-1999;	99US-0142390P
PD	06-SEP-2000.	PR	08-JUL-1999;	99US-0142803P
XX		PR	09-JUL-1999;	99US-0142920P
PF	25-FEB-2000; 2000EP-00301439.	PR	12-JUL-1999;	99US-0142977P
XX		PR	13-JUL-1999;	99US-0143542P
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624P
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005P
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085P
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086P
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144325P
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331P
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332P
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144333P
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334P
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335P
PR	19-APR-1999;	PR	19-JUL-1999;	99US-0144336P
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144337P
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144338P
PR	28-APR-1999;	PR	20-JUL-1999;	99US-0144339P
PR	30-APR-1999;	PR	20-JUL-1999;	99US-0144340P
PR	04-MAY-1999;	PR	20-JUL-1999;	99US-0144884P
PR	05-MAY-1999;	PR	21-JUL-1999;	99US-0144814P
PR	06-MAY-1999;	PR	21-JUL-1999;	99US-0145086P
PR	06-MAY-1999;	PR	21-JUL-1999;	99US-0145088P
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145089P
PR	11-MAY-1999;	PR	22-JUL-1999;	99US-0145089P
PR	14-MAY-1999;	PR	22-JUL-1999;	99US-0145192P
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145145P
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145218P
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224P
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276P
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913P
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145918P
PR	19-MAY-1999;	PR	27-JUL-1999;	99US-0145919P
PR	20-MAY-1999;	PR	28-JUL-1999;	99US-0145951P
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146386P
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389P
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038P
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204P
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302P
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192P
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260P
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303P
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416P
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493P
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935P
PR	14-JUN-1999;	PR	10-AUG-1999;	99US-0148171P
PR	16-JUN-1999;	PR	11-AUG-1999;	99US-0148319P
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341P
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148565P
PR	18-JUN-1999;	PR	13-AUG-1999;	99US-0148684P
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368P
PR	1			

PR 23-AUG-1999; 99US-01499910P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155133P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159323P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160982P.  
PR 22-OCT-1999; 99US-0160983P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.  
Query Match 74.4%; Score 29; DB 3; Length 87;  
Best Local Similarity 83.3%; Pred. No. 2,8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LEHDGI 6  
Db 16 LEHEGI 21  
RESULT 38  
AAE32812  
ID AAE32812 standard; peptide; 92 AA.  
XX  
AC AAE32812;  
XX  
XX 24-MAR-2003 (first entry)  
DT  
XX Human CEA exon #15 encoded peptide.

XX Human; carcinoembryonic antigen; CEA; cell adhesion-mediated disease;  
XX prognosis; cancer; tumour; gene therapy; exon.  
XX Homo sapiens.  
XX WO2002090508-A2.  
XX 14-NOV-2002.  
XX 07-MAY-2002; 2002WO-US014457.  
XX 07-MAY-2001; 2001US-0289179P.  
XX 29-AUG-2001; 2001US-0315736P.  
XX (ALPH-) ALPHAGENE INC.  
XX Stark KA, Weaver A, Hoffmann HM, Krauss R, Valenzuela DB;  
XX Saini KS;  
XX WPI; 2003-111965/10.  
XX N-PSDB; AAD50563.  
XX New carcinoembryonic antigen and polynucleotides encoding them, useful  
XX for treating and/or preventing cell adhesion-mediated disease (e.g.  
XX cancers, solid tumors, tumor metastasis or benign tumors) symptoms.  
XX Claim 17; Page 127; 143pp; English.  
XX The invention relates to novel human carcinoembryonic antigen (CEA)  
XX polypeptides and polynucleotides encoding such polypeptides. CEA genes  
XX are useful as diagnostic and prognostic markers of colon, stomach and  
XX breast cancers. Polypeptides of the invention can be used to diagnose,  
XX treat and/or prevent cell adhesion-mediated disease (e.g. cancers, solid  
XX tumours, tumour metastasis or benign tumors) and to modulate cell-cell  
XX or cell-matrix adhesion in mammalian tissues. The invention is useful in  
XX gene therapy. The present sequence is human CEA exon encoded peptide  
XX Sequence 92 AA;  
Query Match 74.4%; Score 29; DB 6; Length 92;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 EHDGI 6  
Db 65 EHDGI 69  
RESULT 39  
AAU52002  
ID AAU52002 standard; protein; 103 AA.  
XX AC AAU52002;  
XX 27-FEB-2002 (first entry)  
DT  
XX Propionibacterium acnes immunogenic protein #12898.  
DE  
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.  
XX Propionibacterium acnes.  
XX WO200181581-A2.  
XX 01-NOV-2001.  
XX 20-APR-2001; 2001WO-US012865.  
XX 21-APR-2000; 2000US-0199047P.  
PR

PR 02-JUN-2000; 2000US-0208841P.  
PR 07-JUL-2000; 2000US-0216747P.  
XX (CORI-) CORIXA CORP.  
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI: 2001-616774/71.  
DR N-PSDB; AAS59553.  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
PS Example 1; SEQ ID NO 13197; 1069pp; English.  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 103 AA;  
Query Match 74.4%; Score 29; DB 4; Length 103;  
Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 HDGIN 7  
Db 2 HDGVN 6  
RESULT 40  
ID AEM48521 standard; protein; 103 AA.  
XX AEM48521;  
AC AEM48521;  
XX 20-OCT-2003 (first entry)  
XX Propionibacterium acnes predicted ORF-encoded polypeptide #13197.  
DE  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX Propionibacterium acnes.  
OS  
XX WO2003033515-A1.  
PN  
XX 24-APR-2003.  
PD  
XX 11-OCT-2002; 2002WO-US032727.  
PF  
XX 15-OCT-2001; 2001US-00978825.  
PR  
XX (CORI-) CORIXA CORP.  
PA

XX Mitcam JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Valiieve-Douglas J;  
XX WPI: 2003-381789/36.  
DR N-PSDB; ACF64482.  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX Example 1; SEQ ID NO 13197; 1481pp; English.  
PS The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
XX encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 103 AA;  
Query Match 74.4%; Score 29; DB 6; Length 103;  
Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 HDGIN 7  
Db 2 HDGVN 6  
RESULT 41  
ID AAU57868 standard; protein; 106 AA.  
XX AAU57868;  
AC AAU57868;  
XX 27-FEB-2002 (first entry)  
DT  
XX Propionibacterium acnes immunogenic protein #18764.  
DE  
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX Propionibacterium acnes.  
OS  
XX WO200181581-A2.  
PN  
XX 01-NOV-2001.  
PD

XX 20-APR-2001; 2001WO-US012865.  
PF  
XX  
XX 21-APR-2000; 2000US-0199047P.  
PR  
XX 02-JUN-2000; 2000US-0208841P.  
PR  
XX 07-JUL-2000; 2000US-0216747P.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
XX WPI; 2001-616774/71.  
DR N-PSDB; AAS95587.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX  
XX Example 1; SEQ ID NO 19063; 169pp; English.  
PS  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,   
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 106 AA;  
Query Match 74.4%; Score 29; DB 4; Length 106;  
Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 HDGIN 7  
Db 6 HDGVN 10  
RESULT 42  
ABG15770  
ID ABG15770 standard; protein; 106 AA.  
XX  
XX AC ABG15770;  
XX  
XX DT 18-FEB-2002 (first entry)  
XX  
XX DE Novel human diagnostic protein #15761.  
XX  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200175067-A2.  
XX  
XX PD 11-OCT-2001.  
XX  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX  
PF

XX 31-MAR-2000; 2000US-00540217.  
PR  
XX 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS79957.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 46129; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 106 AA;  
Query Match 74.4%; Score 29; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 EHDGI 6  
Db 55 EHDGI 59  
RESULT 43  
ABM54387  
ID ABM54387 standard; protein; 106 AA.  
XX  
XX AC ABM54387;  
XX  
XX DT 20-OCT-2003 (first entry)  
XX  
XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #19063.  
XX  
XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX  
XX OS Propionibacterium acnes.  
XX  
XX PN WO2003033515-A1.  
XX  
XX PD 24-APR-2003.  
XX  
XX PF 11-OCT-2002; 2002WO-US032727.  
XX  
XX PF 15-OCT-2001; 2001US-00978825.  
XX  
PF

XX PA (CORI-) CORIXA CORP.  
 XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Valliave-Douglass J;  
 XX DR WPI; 2003-381789/36.  
 DR N-PSDB; ACF64516.  
 XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX Example 1; SEQ ID NO 19063; 1481pp; English.  
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 106 AA;  
 Query Match 74.4%; Score 29; DB 6; Length 106;  
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HDGIN 7  
 DB 6 HDGVN 10  
 |||||  
 RESULT 44  
 ABG27666  
 ID ABG27666 standard; protein; 107 AA.  
 XX AC ABG27666;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #27657.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 PD

XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 XX 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS91853.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 58025; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 107 AA;  
 Query Match 74.4%; Score 29; DB 4; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LEHDG 5  
 DB 31 LEHDG 35  
 |||||  
 RESULT 45  
 ABG93329  
 ID ABG93329 standard; protein; 113 AA.  
 XX AC ABG93329;  
 XX DT 21-NOV-2002 (first entry)  
 XX DE C. albicans BAX-associated protein fragment SEQ ID 616.  
 XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;  
 KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;  
 KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;  
 KW neurodegeneration; cell death.  
 XX OS Candida albicans.  
 XX WO200264766-A2.  
 XX 22-AUG-2002.  
 PD



XX 21-DEC-2001; 2001WO-EP015398.  
 XX  
 XX 22-DEC-2000; 2000EP-00870318.  
 PR 04-JAN-2001; 2001EP-00870002.  
 PR 09-JAN-2001; 2001EP-00870003.  
 XX  
 XX (JANC ) JANSSEN PHARM NV.  
 PA  
 XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;  
 PI  
 XX  
 XX WPI; 2002-667002/71.  
 DR N-PSDB; ABQ76595.  
 DR  
 XX New isolated nucleic acid representing a synthetic BAX-gene, useful as  
 PT medicament for treating, preventing and/or alleviating yeast or fungal  
 PT infections or proliferative disorders, or for preventing apoptosis in  
 PT certain diseases.  
 PT  
 XX  
 XX Claim 36; Fig 2; 344pp; English.  
 PS  
 XX This invention describes a novel nucleic acid representing a synthetic  
 CC Bax gene. The Bax gene of the invention is useful for identifying Bax-  
 CC resistant yeast or fungi, identifying, or obtaining and identifying  
 CC Candida spp. sequences that are differentially expressed in a pathway  
 CC eventually leading to programmed cell death or identifying inhibitors or  
 CC inhibitor sequences of Bax-induced cell death. The products of the  
 CC invention have cytostatic, fungicide, immunosuppressive, virucide and  
 CC vasotropic activity and can be used in vaccines or for gene therapy. The  
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,  
 CC antisense molecules and antibodies are useful as medicaments or in  
 CC preparing a medicament for treating, preventing and/or alleviating  
 CC diseases associated with yeast or fungi or proliferative disorders, such  
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds  
 CC or polypeptides, or the genetically modified organism are useful for  
 CC preparing a medicament for modifying the endogenic flora of humans and  
 CC other mammals. The vaccine is useful for immunising against yeast or  
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,  
 CC ischaemia, diseases related with viral infections or neurodegenerations.  
 CC This sequence represents a polypeptide associated with the Bax gene  
 CC described in the disclosure of the invention  
 XX  
 XX SQ Sequence 113 AA;  
 Query Match 74.4%; Score 29; DB 5; Length 113;  
 Best Local Similarity 66.7%; Pred. No. 3.7e+02;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGIN 7  
 Db ||:|:  
 58 EHEGLN 63  
 Search completed: June 15, 2004, 12:25:20  
 Job time : 56 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2004, 12:22:52 ; Search time 23 Seconds  
(without alignments)  
15,712 Million cell updates/sec

Title: US-09-394-019C-212

Perfect score: 39

Sequence: 1 LEHGDIN 7

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 290705

Minimum DB seq length: 0

Maximum DB seq length: 150

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	30	76.9	109	2	US-08-470-670A-11
2	30	76.9	109	4	US-08-461-511A-11
3	29	74.4	5	4	US-09-521-650-73
4	29	74.4	5	4	US-09-168-888-73
5	29	74.4	83	4	US-09-107-532A-4957
6	29	74.4	86	4	US-09-134-000C-6650
7	29	74.4	142	4	US-09-543-681A-7907
8	28	71.8	116	1	US-08-497-035-4
9	27	69.2	15	3	US-08-836-561-37
10	27	69.2	15	4	US-09-434-123-37
11	27	69.2	33	4	US-09-331-930A-16
12	27	69.2	33	4	US-09-331-930A-18
13	27	69.2	62	4	US-09-107-532A-6499
14	27	69.2	65	4	US-09-543-681A-6247
15	27	69.2	73	4	US-09-331-930A-2
16	27	69.2	73	4	US-09-331-930A-19
17	27	69.2	73	4	US-09-331-930A-20
18	27	69.2	73	4	US-09-331-930A-21
19	27	69.2	125	4	US-09-543-681A-8087
20	27	69.2	131	3	US-08-836-561-25
21	27	69.2	131	4	US-09-434-123-25
22	26	66.7	40	4	US-09-106-568E-131
23	26	66.7	67	4	US-08-938-165A-289
24	26	66.7	68	4	US-09-540-236-2048
25	26	66.7	70	4	US-09-621-976-4677
26	26	66.7	88	4	US-09-621-976-6918
27	26	66.7	109	4	US-09-325-932A-144

Sequence 4815, Ap  
Sequence 5829, Ap  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 76, Appl  
Sequence 76, Appl  
Sequence 38, Appl  
Sequence 38, Appl  
Sequence 25109, A  
Sequence 10461, A  
Sequence 11391, A  
Sequence 71, Appl  
Sequence 78, Appl  
Sequence 71, Appl  
Sequence 78, Appl  
Sequence 42, Appl  
Sequence 42, Appl  
Sequence 42, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-470-670A-11  
; Sequence 11, Application US/08470670A  
; Patent No. 5834309  
; Patent No. 5834309 5710045  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Craig B. B.  
; APPLICANT: Boise, Lawrence H.  
; TITLE OF INVENTION: VERTEBRATE APOPTOSIS GENE:  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,670A  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,448  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Highlander, Steven L.  
; REGISTRATION NUMBER: 37,642  
; REFERENCE/DOCKET NUMBER: ARCD:090--1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 109 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-470-670A-11

Query Match 76.9%; Score 30; DB 2; Length 109;  
Best Local Similarity 71.4%; Pred. No. 38;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHGDIN 7

```
Db      | |||:|
        45 LFHDGVN 51

RESULT 2
US-08-461-511A-11
; Sequence 11 Application US/08461511A
; Patent No. 6303331
; GENERAL INFORMATION:
; APPLICANT: Thompson, Craig B.B.
;           Boise, Lawrence H.
; TITLE OF INVENTION: VERTEBRATE APOPTOSIS GENE COMPOSITIONS
; AND METHODS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,511A
; FILING DATE: 05-Jun-1995
; CLASSIFICATION: UNKNOWN
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-08-461-511A-11

Query Match      76.9%; Score 30; DB 4; Length 109;
Best Local Similarity 71.4%; Pred. No. 38;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 LEHDGIN 7
        | |||:|
        45 LFHDGVN 51

RESULT 3
US-09-521-650-73
; Sequence 73 Application US/09521650
; Patent No. 6335429
; GENERAL INFORMATION:
; APPLICANT: Weber, Eckard
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Keana, John F.W.
; APPLICANT: Drewe, John A.
; APPLICANT: Zhang, Han-Zhong
; TITLE OF INVENTION: No. 6335429el Fluorogenic or Fluorescent Reporter Molecules and
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1735.0290002
; CURRENT APPLICATION NUMBER: US/09/521,650
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 09/168,888

; EARLIER FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 60/061,582
; EARLIER FILING DATE: 1997-10-10
; EARLIER APPLICATION NUMBER: US 09/033,661
; EARLIER FILING DATE: 1998-03-03
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: Peptide
US-09-168-888-73

Query Match      74.4%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LEHDG 5
        | |||
        1 LEHDG 5

Db      1 LEHDG 5
        | |||
        1 LEHDG 5

Query Match      74.4%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LEHDG 5
        | |||
        1 LEHDG 5

Db      1 LEHDG 5
        | |||
        1 LEHDG 5

RESULT 5
US-09-107-532A-4957
; Sequence 4957 Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
```

```
;
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 4957:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 83 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...83
; SEQUENCE DESCRIPTION: SEQ ID NO: 4957:
US-09-107-532A-4957

Query Match 74.4%; Score 29; DB 4; Length 83;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDDGI 6
Db 57 LKHDGI 62

RESULT 6
US-09-134-000C-6650
; Sequence 6650, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6650
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6650
```

```
Query Match 74.4%; Score 29; DB 4; Length 86;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EHDGIN 7
Db 28 EHSGIN 33

RESULT 7
US-09-543-681A-7907
; Sequence 7907, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7907
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7907

Query Match 74.4%; Score 29; DB 4; Length 142;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGI 6
Db 31 EHDGI 35

RESULT 8
US-08-497-025-4
; Sequence 4, Application US/08497025
; Patent No. 5646251
; GENERAL INFORMATION:
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Rivas, Alberto
; APPLICANT: Laus, Reiner
; APPLICANT: Engleman, Edgar G.
; TITLE OF INVENTION: ALLOREACTION-ASSOCIATED ANTIGEN (ARAg):
; TITLE OF INVENTION: A NOVEL MEMBER OF THE IMMUNOGLOBULIN GENE SUPERFAMILY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/497,025
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/149,212
; FILING DATE: 05-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M
; REGISTRATION NUMBER: 30,223
```

```

; REFERENCE/DOCKET NUMBER: 5490A-219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; FEATURE:
; MOLECULE TYPE: peptide
; NAME/KEY: Peptide
; LOCATION: 1..116
; OTHER INFORMATION: /note= "A variable-type
; OTHER INFORMATION: immunoglobulin-type domain of ARAG-h-1."
US-08-497-025-4

Query Match 71.8%; Score 28; DB 1; Length 116;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDCI 6
Db 52 LQHDGL 57

RESULT 9
US-08-836-561-37
; Sequence 37, Application US/08836561
; Patent No. 6018032
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Masamichi
; APPLICANT: FURUYA, Akiko
; APPLICANT: NAKAMURA, Kazuyasu
; APPLICANT: IIDA, Akihiro
; APPLICANT: ANAZAWA, Hideharu
; APPLICANT: HANAI, No. 6018032uo
; APPLICANT: TAKATSU, Kiyoshi
; TITLE OF INVENTION: Antibody Against Human Interleukin-5
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,561
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 232384/95
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence, III, Stanton T
; REGISTRATION NUMBER: 25,736
; REFERENCE/DOCKET NUMBER: 7005-115-999
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-434-122-37

Query Match 69.2%; Score 27; DB 4; Length 15;
Best Local Similarity 42.9%; Pred. No. 16;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDCIN 7
Db 6 VDHNGVN 12

RESULT 10
US-09-434-122-37
; Sequence 37, Application US/09434122
; Patent No. 6538111
; GENERAL INFORMATION:
; APPLICANT: KOIKE, Masamichi
; APPLICANT: FURUYA, Akiko
; APPLICANT: NAKAMURA, Kazuyasu
; APPLICANT: IIDA, Akihiro
; APPLICANT: ANAZAWA, Hideharu
; APPLICANT: HANAI, No. 6538111uo
; APPLICANT: TAKATSU, Kiyoshi
; TITLE OF INVENTION: Antibody Against Human Interleukin-5
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/434,122
; FILING DATE: 05-NO. 6538111-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,561
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: JP 232384/95
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence, III, Stanton T
; REGISTRATION NUMBER: 25,736
; REFERENCE/DOCKET NUMBER: 7005-115-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-434-122-37

Query Match 69.2%; Score 27; DB 4; Length 15;
Best Local Similarity 42.9%; Pred. No. 16;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDCIN 7
Db 6 VDHNGVN 12
```

Db 6 VDHGVN 12

US-09-331-930A-16

Sequence 16, Application US/09331930A

Patent No. 6436670

GENERAL INFORMATION:

APPLICANT: ZIMMET, PAUL Z.

APPLICANT: COLLIER, GREGORY

TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

FILE REFERENCE: 22975-20007.00

CURRENT APPLICATION NUMBER: US/09/331,930A

CURRENT FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: PCT/AU98/00902

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: AU PP0117/97

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: AU PP0323/97

PRIOR FILING DATE: 1997-11-11

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 16

LENGTH: 33

TYPE: PRT

ORGANISM: Psammomys obesus

US-09-331-930A-16

Query Match 69.2%; Score 27; DB 4; Length 33;

Best Local Similarity 80.0%; Pred. No. 41;

Mismatches 1; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 1;

Qy 3 HDGIN 7

Db 23 HDGMN 27

RESULT 12

US-09-331-930A-18

Sequence 18, Application US/09331930A

Patent No. 6436670

GENERAL INFORMATION:

APPLICANT: ZIMMET, PAUL Z.

APPLICANT: COLLIER, GREGORY

TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

FILE REFERENCE: 22975-20007.00

CURRENT APPLICATION NUMBER: US/09/331,930A

CURRENT FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: PCT/AU98/00902

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: AU PP0117/97

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: AU PP0323/97

PRIOR FILING DATE: 1997-11-11

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 18

LENGTH: 33

TYPE: PRT

ORGANISM: Psammomys obesus

US-09-331-930A-18

Query Match 69.2%; Score 27; DB 4; Length 33;

Best Local Similarity 80.0%; Pred. No. 41;

Mismatches 1; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 1;

Qy 3 HDGIN 7

Db 23 HDGMN 27

RESULT 13

US-09-107-532A-6499

Sequence 6499, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Dereke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6499:

SEQUENCE CHARACTERISTICS:

LENGTH: 62 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (B) LOCATION 1..62

SEQUENCE DESCRIPTION: SEQ ID NO: 6499:

US-09-107-532A-6499

Query Match 69.2%; Score 27; DB 4; Length 62;

Best Local Similarity 86.7%; Pred. No. 85;

Mismatches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGIN 7

Db 20 QHNGIN 25

RESULT 14

US-09-543-681A-6247

Sequence 6247, Application US/09543681A

Patent No. 6605709

GENERAL INFORMATION:

APPLICANT: GARY BRETTON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

FILE REFERENCE: 2709.1002-001

CURRENT APPLICATION NUMBER: US/09/543,681A

CURRENT FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344

SEQ ID NO 6247

Db 6 VDHGVN 12

US-09-331-930A-16

Sequence 16, Application US/09331930A

Patent No. 6436670

GENERAL INFORMATION:

APPLICANT: ZIMMET, PAUL Z.

APPLICANT: COLLIER, GREGORY

TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

FILE REFERENCE: 22975-20007.00

CURRENT APPLICATION NUMBER: US/09/331,930A

CURRENT FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: PCT/AU98/00902

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: AU PP0117/97

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: AU PP0323/97

PRIOR FILING DATE: 1997-11-11

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 16

LENGTH: 33

TYPE: PRT

ORGANISM: Psammomys obesus

US-09-331-930A-16

Query Match 69.2%; Score 27; DB 4; Length 33;

Best Local Similarity 80.0%; Pred. No. 41;

Mismatches 1; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 1;

Qy 3 HDGIN 7

Db 23 HDGMN 27

RESULT 12

US-09-331-930A-18

Sequence 18, Application US/09331930A

Patent No. 6436670

GENERAL INFORMATION:

APPLICANT: ZIMMET, PAUL Z.

APPLICANT: COLLIER, GREGORY

TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

FILE REFERENCE: 22975-20007.00

CURRENT APPLICATION NUMBER: US/09/331,930A

CURRENT FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: PCT/AU98/00902

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: AU PP0117/97

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: AU PP0323/97

PRIOR FILING DATE: 1997-11-11

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 18

LENGTH: 33

TYPE: PRT

ORGANISM: Psammomys obesus

US-09-331-930A-18

Query Match 69.2%; Score 27; DB 4; Length 33;

Best Local Similarity 80.0%; Pred. No. 41;

Mismatches 1; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 1;

Qy 3 HDGIN 7

Db 23 HDGMN 27

RESULT 13

US-09-107-532A-6499

Sequence 6499, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Dereke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6499:

SEQUENCE CHARACTERISTICS:

LENGTH: 62 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (B) LOCATION 1..62

SEQUENCE DESCRIPTION: SEQ ID NO: 6499:

US-09-107-532A-6499

Query Match 69.2%; Score 27; DB 4; Length 62;

Best Local Similarity 86.7%; Pred. No. 85;

Mismatches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGIN 7

Db 20 QHNGIN 25

RESULT 14

US-09-543-681A-6247

Sequence 6247, Application US/09543681A

Patent No. 6605709

GENERAL INFORMATION:

APPLICANT: GARY BRETTON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

FILE REFERENCE: 2709.1002-001

CURRENT APPLICATION NUMBER: US/09/543,681A

CURRENT FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344

SEQ ID NO 6247

```
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6247

Query Match          69.2%; Score 27; DB 4; Length 65;
Best Local Similarity 57.1%; Pred. No. 89;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LEHGIN 7
   |||:|
Db 25 LSHGVN 31

RESULT 15
US-09-331-930A-2
; Sequence 2, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.
; APPLICANT: COLLIER, GREGORY
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-20007.00
; CURRENT APPLICATION NUMBER: US/09/331,930A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Amino acid
; OTHER INFORMATION: sequence for beacon from unknown organism
US-09-331-930A-2

Query Match          69.2%; Score 27; DB 4; Length 73;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7
   |||:|
Db 63 HDGMN 67

RESULT 16
US-09-331-930A-19
; Sequence 19, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.
; APPLICANT: COLLIER, GREGORY
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-20007.00
; CURRENT APPLICATION NUMBER: US/09/331,930A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Murine sp.
US-09-331-930A-21

Query Match          69.2%; Score 27; DB 4; Length 73;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7
   |||:|
Db 63 HDGMN 67

RESULT 17
US-09-331-930A-20
; Sequence 20, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.
; APPLICANT: COLLIER, GREGORY
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-20007.00
; CURRENT APPLICATION NUMBER: US/09/331,930A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-930A-20

Query Match          69.2%; Score 27; DB 4; Length 73;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7
   |||:|
Db 63 HDGMN 67

RESULT 18
US-09-331-930A-21
; Sequence 21, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.
; APPLICANT: COLLIER, GREGORY
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-20007.00
; CURRENT APPLICATION NUMBER: US/09/331,930A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Murine sp.
US-09-331-930A-21
```

Query Match 69.2%; Score 27; DB 4; Length 73;  
Best Local Similarity 80.0%; Pred. No. 1e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
Db 63 HDGMN 67

US-09-543-681A-8087  
Sequence 8087, Application US/09543681A  
Patent No. 6505709  
GENERAL INFORMATION:  
APPLICANT: GARY BRSTON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 8087  
LENGTH: 125  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-8087

Query Match 69.2%; Score 27; DB 4; Length 125;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDCI 6  
Db 119 LEHGV 124

US-08-836-561-25  
Sequence 25, Application US/08836561  
Patent No. 6018032  
GENERAL INFORMATION:  
APPLICANT: KOIKE, Masamichi  
APPLICANT: FURUYA, Akiko  
APPLICANT: NAKAMURA, Kazuyasu  
APPLICANT: IIDA, Akihiro  
APPLICANT: ANAZAWA, Hideharu  
APPLICANT: HANAI, No. 6018032uo  
APPLICANT: TAKATSU, Kiyoshi  
TITLE OF INVENTION: Antibody Against Human Interleukin-5  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,561  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 232384/95  
FILING DATE: 11-SEP-1995

Query Match 69.2%; Score 27; DB 3; Length 131;  
Best Local Similarity 42.9%; Pred. No. 2e+02;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDCIN 7  
Db 49 VDHNGVN 55

US-08-836-561-25  
Sequence 25, Application US/09434122  
Patent No. 6538111  
GENERAL INFORMATION:  
APPLICANT: KOIKE, Masamichi  
APPLICANT: FURUYA, Akiko  
APPLICANT: NAKAMURA, Kazuyasu  
APPLICANT: IIDA, Akihiro  
APPLICANT: ANAZAWA, Hideharu  
APPLICANT: HANAI, No. 6538111uo  
APPLICANT: TAKATSU, Kiyoshi  
TITLE OF INVENTION: Antibody Against Human Interleukin-5  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/434,122  
FILING DATE: 05-NO. 6538111-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,561  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: JP 232384/95  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lawrence, III, Stanton T  
REGISTRATION NUMBER: 25,736  
REFERENCE/DOCKET NUMBER: 7005-115-999  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 amino acids  
TYPE: amino acid

ATTORNEY/AGENT INFORMATION:  
NAME: Lawrence, III, Stanton T  
REGISTRATION NUMBER: 25,736  
REFERENCE/DOCKET NUMBER: 7005-115-999  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 amino acids  
TYPE: amino acid  
MOLUCULE TYPE: peptide  
US-08-836-561-25

Query Match 69.2%; Score 27; DB 3; Length 131;  
Best Local Similarity 42.9%; Pred. No. 2e+02;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDCIN 7  
Db 49 VDHNGVN 55

US-08-836-561-25  
Sequence 25, Application US/09434122  
Patent No. 6538111  
GENERAL INFORMATION:  
APPLICANT: KOIKE, Masamichi  
APPLICANT: FURUYA, Akiko  
APPLICANT: NAKAMURA, Kazuyasu  
APPLICANT: IIDA, Akihiro  
APPLICANT: ANAZAWA, Hideharu  
APPLICANT: HANAI, No. 6538111uo  
APPLICANT: TAKATSU, Kiyoshi  
TITLE OF INVENTION: Antibody Against Human Interleukin-5  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/434,122  
FILING DATE: 05-NO. 6538111-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,561  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: JP 232384/95  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lawrence, III, Stanton T  
REGISTRATION NUMBER: 25,736  
REFERENCE/DOCKET NUMBER: 7005-115-999  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 amino acids  
TYPE: amino acid



STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
US-09-434-122-25

Query Match 69.2%; Score 27; DB 4; Length 131;  
Best Local Similarity 42.9%; Pred. No. 2e+02;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHNGIN 7  
Db 49 VDHNGVN 55

RESULT 22  
US-09-106-568E-131  
; Sequence 131, Application US/09106568E  
; Patent No. 6455248  
; GENERAL INFORMATION:  
; APPLICANT: Bhattacharjee, J.  
; APPLICANT: Bhattacharjee, Kalavati  
; APPLICANT: Bhattacharjee, Vasker  
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN  
; FILE REFERENCE: 96,247-A  
; CURRENT APPLICATION NUMBER: US/09/106,568E  
; CURRENT FILING DATE: 1998-06-29  
; PRIOR APPLICATION NUMBER: 08/650,809  
; PRIOR FILING DATE: 1997-05-20  
; NUMBER OF SEQ ID NOS: 160  
; SOFTWARE: Microsoft Word 97  
; SEQ ID NO 131  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Polypeptide segment of TYCA\_BACER shown in Figure 4.  
US-09-106-568E-131

Query Match 66.7%; Score 26; DB 4; Length 40;  
Best Local Similarity 83.3%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHNGI 6  
Db 22 LEHNGI 27

RESULT 23  
US-08-936-165A-289  
; Sequence 289, Application US/08936165A  
; Patent No. 6348582  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Burnham, Martin  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Lonetto, Michael  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Pratt, Julie  
; APPLICANT: Reichard, Richard  
; APPLICANT: Rosenberg, Martin  
; APPLICANT: Ward, Judith  
; TITLE OF INVENTION: NO. 6348582el Prokaryotic Polynucleotides,  
; TITLE OF INVENTION: Polypeptides and Their Uses  
; NUMBER OF SEQUENCES: 534  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA

ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 289:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 67 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-936-165A-289

Query Match 66.7%; Score 26; DB 4; Length 67;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHNGI 6  
Db 17 LEYDGV 22

RESULT 24  
US-09-540-236-2048  
; Sequence 2048, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRH  
; FILE REFERENCE: 2709,2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 2048  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: M.catarrhalis  
US-09-540-236-2048

Query Match 66.7%; Score 26; DB 4; Length 68;  
Best Local Similarity 57.1%; Pred. No. 1.5e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHNGIN 7  
Db 51 IKNDGIN 57

RESULT 25  
US-09-621-976-4677  
; Sequence 4677, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 4677  
; LENGTH: 70  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-4677

Query Match 66.7%; Score 26; DB 4; Length 70;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDS 5  
Db 48 LQHDG 52

RESULT 26  
US-09-621-976-6918  
; Sequence 6918, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 6918  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-6918

Query Match 66.7%; Score 26; DB 4; Length 88;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDS 6  
Db 41 LEYDGV 46

RESULT 27  
US-09-325-932A-144  
; Sequence 144, Application US/09325932A  
; Patent No. 6451604  
; GENERAL INFORMATION:  
; APPLICANT: Flinn, Barry  
; APPLICANT: Lasham, Annette  
; TITLE OF INVENTION: Compositions affecting programmed cell  
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop  
; FILE REFERENCE: 1022  
; CURRENT APPLICATION NUMBER: US/09/325,932A  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 206  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 144  
; LENGTH: 109  
; TYPE: PRT  
; ORGANISM: Eucalyptus grandis  
US-09-325-932A-144

Query Match 66.7%; Score 26; DB 4; Length 109;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 EHDGIN 7  
Db 81 EHTGLN 86

RESULT 28  
US-09-328-352-4815  
; Sequence 4815, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 3252  
; SEQ ID NO 4815  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4815

Query Match 66.7%; Score 26; DB 4; Length 113;  
Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDS 6  
Db 51 IKHDGV 56

RESULT 29  
US-09-621-976-5829  
; Sequence 5829, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5829  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -88..-1  
US-09-621-976-5829

Query Match 66.7%; Score 26; DB 4; Length 113;  
Best Local Similarity 57.1%; Pred. No. 2.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDS 7  
Db 64 LEHDDVS 70

RESULT 30  
US-08-047-033-2  
; Sequence 2, Application US/08047033  
; Patent No. 544157  
; GENERAL INFORMATION:  
; APPLICANT: Fujio SUZUKI et al.  
; TITLE OF INVENTION: CHONDROMODULIN-I PROTEIN

```

/  AUTHORS:
/  TITLE:
/  JOURNAL:
/  VOLUME:
/  ISSUE:
/  PAGES:
/  DATE:
/  DOCUMENT NUMBER:
/  FILING DATE:
/  PUBLICATION DATE:
/  RELEVANT RESIDUES IN SEQ ID NO:
/
US-08-047-033-2
/
Query Match      66.7%; Score 26; DB 1; Length 120;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels
/
QY      1 LEHDGI 6
        |:|:|
Db      62 LDHGI 67
/
RESULT 31
US-08-047-033-1
; Sequence 1, Application US/08047033
; Patent No. 5444157
; GENERAL INFORMATION:
; APPLICANT: Fujio SUZUKI et al.
; TITLE OF INVENTION: CHONDROMODULIN-I PROTEIN
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/047,033
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/821,859
; FILING DATE:
; APPLICATION NUMBER: US/07/745,497
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 121 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: bovine
; STRAIN:
; INDIVIDUAL ISOLATE:

```

DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE: fetal cartilage  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: modified site  
LOCATION: 7  
IDENTIFICATION METHOD:  
OTHER INFORMATION: /note= "Xaa is Met or Val"  
FEATURE:  
NAME/KEY: modified site  
LOCATION: 10  
IDENTIFICATION METHOD:  
OTHER INFORMATION: /note= "Xaa is Glu or Thr"  
FEATURE:  
NAME/KEY: modified site  
LOCATION: 84  
IDENTIFICATION METHOD:  
OTHER INFORMATION: /note= "Xaa is Cys or Val"  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-047-033-1

Query Match 66.7%; Score 26; DB 1; Length 121;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDCI 6  
|:|:|  
Db 63 LDHEGI 68

RESULT 32  
US-08-887-534A-76  
; Sequence 76, Application US/08887534A  
; Patent No. 6455323  
; GENERAL INFORMATION:  
; APPLICANT: Holden, David W.  
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive/6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/887,534  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/887,534  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 28341/33996  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: (312) 474-6600  
; INFORMATION FOR SEQ ID NO: 76:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 128 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-527-431-76

Query Match 66.7%; Score 26; DB 4; Length 128;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 28341/33996  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: (312) 474-6600  
INFORMATION FOR SEQ ID NO: 76:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 128 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-887-534A-76

Query Match 66.7%; Score 26; DB 4; Length 128;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|:|:|  
Db 39 LQHDG 43

RESULT 33  
US-09-527-431-76  
; Sequence 76, Application US/09527431  
; Patent No. 6485899  
; GENERAL INFORMATION:  
; APPLICANT: Holden, David W.  
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive/6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/527,431  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/887,534  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 28341/33996  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: (312) 474-6600  
; INFORMATION FOR SEQ ID NO: 76:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 128 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-527-431-76

Query Match 66.7%; Score 26; DB 4; Length 128;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDG 5  
Db 39 LQHDG 43

RESULT 34  
US-08-392-828C-38  
; Sequence 38 Application US/08392828C  
; Patent No. 5795962  
; GENERAL INFORMATION:  
; APPLICANT: IWANAGA, SADAOKI  
; APPLICANT: MUTA, TATSUSHI  
; APPLICANT: SEKI, NORIAKI  
; APPLICANT: ODA, TOSHIO  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE AND DNA ENCODING  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &  
; ADDRESSEE: THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/392,828C  
; FILING DATE: 28-FEB-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CAMPBELL, PAULA A.  
; REGISTRATION NUMBER: 32,503  
; REFERENCE/DOCKET NUMBER: FUN-033  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 137 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..137  
; OTHER INFORMATION: /note= "XYN Z SEQUENCE (FIGURE 3)"  
US-08-392-828C-38

Query Match 66.7%; Score 26; DB 1; Length 137;  
Best Local Similarity 66.7%; Pred. No. 3.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGIN 7  
Db 8 DYDGIN 13

RESULT 35  
US-09-330-945-38  
; Sequence 38 Application US/09330945  
; Patent No. 6077946  
; GENERAL INFORMATION:  
; APPLICANT: IWANAGA, SADAOKI  
; APPLICANT: MUTA, TATSUSHI  
; APPLICANT: SEKI, NORIAKI  
; APPLICANT: ODA, TOSHIO

; TITLE OF INVENTION: DNA ENCODING HORSESHOE CRAB  
; TITLE OF INVENTION: AMEBOCYTE LYSATE FACTOR G SUBUNIT A  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &  
; ADDRESSEE: THIBEAULT, LLP  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/330,945  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: 09/119,995  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER, EDMUND R.  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: FUN-032DV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 137 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..137  
; OTHER INFORMATION: /note= "XYN Z SEQUENCE (FIGURE 3)"  
US-09-330-945-38

Query Match 66.7%; Score 26; DB 3; Length 137;  
Best Local Similarity 66.7%; Pred. No. 3.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGIN 7  
Db 8 DYDGIN 13

RESULT 36  
US-09-252-991A-25109  
; Sequence 25109, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25109  
; TYPE: PRT  
; LENGTH: 139  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25109

Query Match 66.7%; Score 26; DB 4; Length 139;  
Best Local Similarity 83.3%; Pred. No. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
|  
|  
|  
|  
|  
82 LVHDOI 87

Db

RESULT 37  
US-09-489-039A-10461  
; Sequence 10461, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 10461  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
; US-09-489-039A-10461

Query Match 66.7%; Score 26; DB 4; Length 142;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
:|:|:|  
15 IEHNGI 20

Db

RESULT 38  
US-09-489-039A-11391  
; Sequence 11391, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 11391  
; LENGTH: 147  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
; US-09-489-039A-11391

Query Match 66.7%; Score 26; DB 4; Length 147;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
:|:|:|  
46 IDHDL 51

Db

RESULT 39  
US-09-521-650-71  
; Sequence 71, Application US/09521650  
; Patent No. 6335429  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Eckard  
; APPLICANT: Cai, Sui Xiong  
; APPLICANT: Keana, John F.W.  
; APPLICANT: Drewe, John A.  
; APPLICANT: Zhang, Han-Zhong  
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence  
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the  
; TITLE OF INVENTION: Use Thereof  
; FILE REFERENCE: 1735.0290002  
; CURRENT APPLICATION NUMBER: US/09/521,650  
; CURRENT FILING DATE: 2000-03-08  
; EARLIER FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 60/061,582  
; EARLIER FILING DATE: 1997-10-10  
; EARLIER APPLICATION NUMBER: US 09/033,661  
; EARLIER FILING DATE: 1998-03-03  
; NUMBER OF SEQ ID NOS: 142  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 71  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic  
; OTHER INFORMATION: Peptide  
US-09-521-650-71

Query Match 64.1%; Score 25; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDG 5  
|  
|  
|  
|  
|  
2 EHDG 5

Db

RESULT 40  
US-09-521-650-78  
; Sequence 78, Application US/09521650  
; Patent No. 6335429  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Eckard  
; APPLICANT: Cai, Sui Xiong  
; APPLICANT: Keana, John F.W.  
; APPLICANT: Drewe, John A.  
; APPLICANT: Zhang, Han-Zhong  
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence  
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the  
; TITLE OF INVENTION: Use Thereof  
; FILE REFERENCE: 1735.0290002  
; CURRENT APPLICATION NUMBER: US/09/521,650  
; CURRENT FILING DATE: 2000-03-08  
; EARLIER FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 60/061,582  
; EARLIER FILING DATE: 1997-10-10  
; EARLIER APPLICATION NUMBER: US 09/033,661  
; EARLIER FILING DATE: 1998-03-03  
; NUMBER OF SEQ ID NOS: 142  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 78  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic  
; OTHER INFORMATION: Peptide  
US-09-521-650-78

Query Match 64.1%; Score 25; DB 4; Length 5;

Query Match 66.7%; Score 26; DB 4; Length 139;  
Best Local Similarity 83.3%; Pred. No. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
|  
|  
|  
|  
|  
82 LVHDOI 87

Db

RESULT 37  
US-09-489-039A-10461  
; Sequence 10461, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 10461  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
; US-09-489-039A-10461

Query Match 66.7%; Score 26; DB 4; Length 142;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
:|:|:|  
15 IEHNGI 20

Db

RESULT 38  
US-09-489-039A-11391  
; Sequence 11391, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 11391  
; LENGTH: 147  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
; US-09-489-039A-11391

Query Match 66.7%; Score 26; DB 4; Length 147;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
:|:|:|  
46 IDHDL 51

Db

RESULT 39  
US-09-521-650-71  
; Sequence 71, Application US/09521650  
; Patent No. 6335429  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Eckard  
; APPLICANT: Cai, Sui Xiong  
; APPLICANT: Keana, John F.W.  
; APPLICANT: Drewe, John A.  
; APPLICANT: Zhang, Han-Zhong  
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence  
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the  
; TITLE OF INVENTION: Use Thereof  
; FILE REFERENCE: 1735.0290002  
; CURRENT APPLICATION NUMBER: US/09/521,650  
; CURRENT FILING DATE: 2000-03-08  
; EARLIER FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 60/061,582  
; EARLIER FILING DATE: 1997-10-10  
; EARLIER APPLICATION NUMBER: US 09/033,661  
; EARLIER FILING DATE: 1998-03-03  
; NUMBER OF SEQ ID NOS: 142  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 71  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic  
; OTHER INFORMATION: Peptide  
US-09-521-650-71

Query Match 64.1%; Score 25; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDG 5  
|  
|  
|  
|  
|  
2 EHDG 5

Db

RESULT 40  
US-09-521-650-78  
; Sequence 78, Application US/09521650  
; Patent No. 6335429  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Eckard  
; APPLICANT: Cai, Sui Xiong  
; APPLICANT: Keana, John F.W.  
; APPLICANT: Drewe, John A.  
; APPLICANT: Zhang, Han-Zhong  
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence  
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the  
; TITLE OF INVENTION: Use Thereof  
; FILE REFERENCE: 1735.0290002  
; CURRENT APPLICATION NUMBER: US/09/521,650  
; CURRENT FILING DATE: 2000-03-08  
; EARLIER FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 60/061,582  
; EARLIER FILING DATE: 1997-10-10  
; EARLIER APPLICATION NUMBER: US 09/033,661  
; EARLIER FILING DATE: 1998-03-03  
; NUMBER OF SEQ ID NOS: 142  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 78  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic  
; OTHER INFORMATION: Peptide  
US-09-521-650-78

Query Match 64.1%; Score 25; DB 4; Length 5;

```

Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 EHDG 5
      |||
Db      2 EHDG 5

```

```

RESULT 41
US-09-168-888-71
; Sequence 71, Application US/09168888
; Patent No. 6342611
; GENERAL INFORMATION:
; APPLICANT: Weber, Eckard
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Keana, John F.W.
; APPLICANT: Drewe, John A.
; APPLICANT: Zhang, Han-Zhong
; TITLE OF INVENTION: No. 6342611el Fluorogenic or Fluorescent Reporter Molecules and
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
; TITLE OF INVENTION: Use thereof
; FILE REFERENCE: 1735.0290002
; CURRENT APPLICATION NUMBER: US/09/168.888
; CURRENT FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 60/061,582
; EARLIER FILING DATE: 1997-10-10
; EARLIER APPLICATION NUMBER: US 09/033,661
; EARLIER FILING DATE: 1998-03-03
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-168-888-71

```

```

Query Match      64.1%; Score 25; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 EHDG 5
Db      2 EHDG 5

```

RESULT 42  
 US-09-168-888-78  
 ; Sequence 78, Application US/09168888  
 ; Patent No. 6342611  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weber, Eckard  
 ; APPLICANT: Cai, Sui Xiong  
 ; APPLICANT: Keana, John F.W.  
 ; APPLICANT: Drewe, John A.  
 ; APPLICANT: Zhang, Han-Zhong  
 ; TITLE OF INVENTION: No. 6342611e1 Fluorogenic or Fluorescent Reporter Molecules and  
 ; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence  
 ; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the  
 ; TITLE OF INVENTION: Use thereof  
 ; FILE REFERENCE: 1735.0290002  
 ; CURRENT APPLICATION NUMBER: US/09/168,888  
 ; CURRENT FILING DATE: 1998-10-09  
 ; EARLIER APPLICATION NUMBER: US 60/061,582  
 ; EARLIER FILING DATE: 1997-10-10  
 ; EARLIER APPLICATION NUMBER: US 09/033,661  
 ; EARLIER FILING DATE: 1998-03-03  
 ; NUMBER OF SEQ ID NOS: 142  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 78

Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 EHDG 5  
|||  
Db 2 EHDG 5

RESULT 45

US-09-954-697-42  
; Sequence 42, Application US/09954697  
; Patent No. 6610541  
; GENERAL INFORMATION:  
; APPLICANT: Alnemri, Emad S.  
; TITLE OF INVENTION: RECOMBINANT, ACTIVE CASPASES AND USSES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 480140.431D2  
; CURRENT APPLICATION NUMBER: US/09/954,697  
; CURRENT FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 42  
; LENGTH: 5  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Protease sensitive linker region between the small  
; OTHER INFORMATION: and large subunits of a Rev-caspase  
US-09-954-697-42

Query Match 64.1%; Score 25; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 EHDG 5  
|||  
Db 2 EHDG 5

Search completed: June 15, 2004, 12:26:30  
Job time : 24 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2004, 12:23:26 ; Search time 42 seconds  
(without alignments)  
46.954 Million cell updates/sec

Title: US-09-394-019c-212

Perfect score: 39

Sequence: 1 LEHDGIN 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1158786 seqs, 281726120 residues

Total number of hits satisfying chosen parameters: 581086

Minimum DB seq length: 0

Maximum DB seq length: 150

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA: \*

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	18	US-09-747-287-199	Sequence 199, App
2	39	100.0	18	US-09-874-350A-165	Sequence 165, App
3	39	100.0	18	US-09-874-350A-206	Sequence 206, App
4	35	89.7	18	US-09-747-287-201	Sequence 201, App
5	35	89.7	18	US-09-874-350A-167	Sequence 167, App
6	35	89.7	18	US-09-874-350A-207	Sequence 207, App
7	35	89.7	131	US-10-424-599-170453	Sequence 170453, App
8	33	84.6	76	US-10-424-599-284592	Sequence 284592, App
9	32	82.1	17	US-10-083-357-1105	Sequence 1105, App
10	31	79.5	30	US-09-270-983-4	Sequence 4, Appl
11	31	79.5	66	US-10-424-599-234026	Sequence 234026, App
12	30	76.9	18	US-09-747-287-203	Sequence 203, App
13	30	76.9	18	US-09-874-350A-169	Sequence 169, App
14	30	76.9	55	US-10-424-599-210542	Sequence 210542, App
15	30	76.9	70	US-10-424-599-160368	Sequence 160368, App

16	30	76.9	121	12	US-10-424-599-221259	Sequence 221259, App
17	30	76.9	128	12	US-10-425-114-70959	Sequence 70959, App
18	29	74.4	5	9	US-09-747-387-73	Sequence 73, Appl
19	29	74.4	18	12	US-09-747-287-200	Sequence 200, App
20	29	74.4	18	12	US-09-874-350A-166	Sequence 166, App
21	29	74.4	20	10	US-08-853-079-215	Sequence 215, App
22	29	74.4	20	16	US-10-254-443-215	Sequence 215, App
23	29	74.4	84	9	US-09-867-550-774	Sequence 774, App
24	29	74.4	98	12	US-10-424-599-165583	Sequence 165583, App
25	29	74.4	101	12	US-10-424-599-203990	Sequence 203990, App
26	29	74.4	127	15	US-10-369-493-11679	Sequence 11679, App
27	29	74.4	127	15	US-10-369-493-14258	Sequence 14258, App
28	29	74.4	132	15	US-10-369-493-14694	Sequence 14694, App
29	29	74.4	137	14	US-10-424-599-147562	Sequence 147562, App
30	29	74.4	137	14	US-10-198-070-26	Sequence 26, Appl
31	29	74.4	137	16	US-10-422-571-19	Sequence 19, Appl
32	29	74.4	137	16	US-10-422-571-29	Sequence 29, Appl
33	29	74.4	137	16	US-10-422-571-38	Sequence 38, Appl
34	29	74.4	137	16	US-10-422-571-39	Sequence 39, Appl
35	29	74.4	137	16	US-10-422-571-89	Sequence 89, Appl
36	29	74.4	137	16	US-10-422-571-126	Sequence 126, App
37	29	74.4	137	16	US-10-422-571-130	Sequence 130, App
38	29	74.4	143	12	US-10-221-278-647	Sequence 647, App
39	29	74.4	143	15	US-10-291-172-647	Sequence 647, App
40	28	71.8	49	12	US-10-424-599-262285	Sequence 262285, App
41	28	71.8	52	11	US-09-864-408A-116	Sequence 116, App
42	28	71.8	55	12	US-10-424-599-208279	Sequence 208279, App
43	28	71.8	55	12	US-10-424-599-222247	Sequence 222247, App
44	28	71.8	68	12	US-10-424-599-215145	Sequence 215145, App
45	28	71.8	70	12	US-10-424-599-216092	Sequence 216092, App

#### ALIGNMENTS

#### RESULT 1

US-09-747-287-199  
Sequence 199, Application US/09747287  
Publication No. US20030207264A1

GENERAL INFORMATION:  
APPLICANT: KOMORIYA, AKIRA

TITLE OF INVENTION: HOMO-DOUBLY LABELED COMPOSITIONS FOR THE DETECTION OF ENZYME  
TITLE OF INVENTION: ACTIVITY IN BIOLOGICAL SAMPLES

FILE REFERENCE: 300-948600US  
CURRENT APPLICATION NUMBER: US/09/747,287

CURRENT FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: US 09/349,019

PRIOR FILING DATE: 1999-09-10

PRIOR APPLICATION NUMBER: US08/802,981

PRIOR FILING DATE: 1997-02-20

PRIOR APPLICATION NUMBER: PCT/US00/24882

PRIOR FILING DATE: 2000-09-11

NUMBER OF SEQ ID NOS: 242

SOFTWARE: PatentIn version 3.2

SEQ ID NO 199

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial

FEATURE:

OTHER INFORMATION: Synthetic peptide.

NAME/KEY: misc feature

LOCATION: (4)..(4)

OTHER INFORMATION: Xaa is epsilon-aminocaproic acid

NAME/KEY: misc feature

LOCATION: (14)..(14)

OTHER INFORMATION: Xaa is epsilon-aminocaproic acid

US-09-747-287-199

Query Match 100.0%; Score 39; DB 12; Length 18;

Best Local Similarity 100.0%; Pred. No. 0.38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LEHDGIN 7  
| | | | |  
Db 6 LEHDGIN 12

## RESULT 2

US-09-874-350A-165  
; Sequence 165, Application US/09874350A  
; Publication No. US20040096926A1  
; GENERAL INFORMATION:  
; APPLICANT: Oncoimmunin, Inc.  
; APPLICANT: Packard, Beverly  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL  
; TITLE OF INVENTION: AND METHODS OF USE THEREOF  
; FILE REFERENCE: 300-903840US  
; CURRENT APPLICATION NUMBER: US/09/874,350A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: PCT/US98/00300  
; PRIOR FILING DATE: 1998-02-20  
; PRIOR APPLICATION NUMBER: PCT/US00/24882  
; PRIOR FILING DATE: 2000-09-11  
; PRIOR APPLICATION NUMBER: US 09/394,019  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: US 08/802,981  
; PRIOR FILING DATE: 1997-02-20  
; NUMBER OF SEQ ID NOS: 221  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 165  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: K is blocked with Fmoc  
; NAME/KEY: MOD RES  
; LOCATION: (14)..(14)  
; OTHER INFORMATION: X is epsilon-aminocaproic acid  
; NAME/KEY: MOD RES  
; LOCATION: (4)..(4)  
; OTHER INFORMATION: X is epsilon-aminocaproic acid  
US-09-874-350A-165

Query Match 100.0%; Score 39; DB 12; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
| | | | |  
Db 6 LEHDGIN 12

## RESULT 3

US-09-874-350A-206  
; Sequence 206, Application US/09874350A  
; Publication No. US20040096926A1  
; GENERAL INFORMATION:  
; APPLICANT: Oncoimmunin, Inc.  
; APPLICANT: Packard, Beverly  
; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL  
; TITLE OF INVENTION: AND METHODS OF USE THEREOF  
; FILE REFERENCE: 300-903840US  
; CURRENT APPLICATION NUMBER: US/09/874,350A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: PCT/US98/00300  
; PRIOR FILING DATE: 1998-02-20  
; PRIOR APPLICATION NUMBER: PCT/US00/24882  
; PRIOR FILING DATE: 2000-09-11

; PRIOR APPLICATION NUMBER: US 09/394,019  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: US 08/802,981  
; PRIOR FILING DATE: 1997-02-20  
; NUMBER OF SEQ ID NOS: 221  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 206  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Protease indicator  
; NAME/KEY: misc\_feature  
; LOCATION: (4)..(4)  
; OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
; NAME/KEY: misc\_feature  
; LOCATION: (14)..(14)  
; OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
US-09-874-350A-206

Query Match 100.0%; Score 39; DB 12; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
| | | | |  
Db 6 LEHDGIN 12

## RESULT 4

US-09-747-287-201  
; Sequence 201, Application US/09747287  
; Publication No. US20030207264A1  
; GENERAL INFORMATION:  
; APPLICANT: KOMORIYA, AKIRA  
; TITLE OF INVENTION: HOMO-DOUBLY LABELED COMPOSITIONS FOR THE DETECTION OF ENZYME  
; TITLE OF INVENTION: ACTIVITY IN BIOLOGICAL SAMPLES  
; FILE REFERENCE: 300-948600US  
; CURRENT APPLICATION NUMBER: US/09/747,287  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/349,019  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: US08/802,981  
; PRIOR FILING DATE: 1997-02-20  
; PRIOR APPLICATION NUMBER: PCT/US00/24882  
; PRIOR FILING DATE: 2000-09-11  
; NUMBER OF SEQ ID NOS: 242  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 201  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Synthetic peptide.  
; NAME/KEY: misc\_feature  
; LOCATION: (4)..(4)  
; OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
; NAME/KEY: misc\_feature  
; LOCATION: (14)..(14)  
; OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
US-09-747-287-201

Query Match 89.7%; Score 35; DB 12; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
| | | | |  
Db 7 EHDGIN 12

RESULT 5  
 US-09-874-350A-167  
 ; Sequence 167, Application US/09874350A  
 ; Publication No. US2004009626A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oncoimmunin, Inc.  
 ; APPLICANT: Komoriya, Akira  
 ; APPLICANT: Packard, Beverly  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL S  
 ; FILE REFERENCE: 300-903840US  
 ; CURRENT APPLICATION NUMBER: US/09/874,350A  
 ; CURRENT FILING DATE: 2001-06-04  
 ; PRIOR APPLICATION NUMBER: PCT/US98/00300  
 ; PRIOR FILING DATE: 1998-02-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/24882  
 ; PRIOR FILING DATE: 2000-09-11  
 ; PRIOR APPLICATION NUMBER: US 09/394,019  
 ; PRIOR FILING DATE: 1999-09-10  
 ; PRIOR APPLICATION NUMBER: US 08/802,981  
 ; PRIOR FILING DATE: 1997-02-20  
 ; NUMBER OF SEQ ID NOS: 221  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 167  
 ; LENGTH: 18  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; OTHER INFORMATION: Synthetic peptide substrate  
 ; NAME/KEY: MOD\_RES  
 ; LOCATION: (1)..(1)  
 ; OTHER INFORMATION: K is blocked with Fmoc  
 ; NAME/KEY: MOD\_RES  
 ; LOCATION: (14)..(14)  
 ; OTHER INFORMATION: X is epsilon-aminocaproic acid  
 ; NAME/KEY: MOD\_RES  
 ; LOCATION: (4)..(4)  
 ; OTHER INFORMATION: X is epsilon aminocaproic acid  
 US-09-874-350A-167  
 Query Match 89.7%; Score 35; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 EHDGIN 7  
 Db 7 EHDGIN 12  
 RESULT 6  
 US-09-874-350A-207  
 ; Sequence 207, Application US/09874350A  
 ; Publication No. US2004009626A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oncoimmunin, Inc.  
 ; APPLICANT: Komoriya, Akira  
 ; APPLICANT: Packard, Beverly  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL S  
 ; FILE REFERENCE: 300-903840US  
 ; CURRENT APPLICATION NUMBER: US/09/874,350A  
 ; CURRENT FILING DATE: 2001-06-04  
 ; PRIOR APPLICATION NUMBER: PCT/US98/00300  
 ; PRIOR FILING DATE: 1998-02-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/24882  
 ; PRIOR FILING DATE: 2000-09-11  
 ; PRIOR APPLICATION NUMBER: US 09/394,019  
 ; PRIOR FILING DATE: 1999-09-10  
 ; PRIOR APPLICATION NUMBER: US 08/802,981  
 ; PRIOR FILING DATE: 1997-02-20  
 ; NUMBER OF SEQ ID NOS: 221  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 207  
 ; LENGTH: 18  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; OTHER INFORMATION: Protease indicator  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4)..(4)  
 ; OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (14)..(14)  
 ; OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
 US-09-874-350A-207  
 Query Match 89.7%; Score 35; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 EHDGIN 7  
 Db 106 EHDGIN 111  
 RESULT 7  
 US-10-424-599-170453  
 ; Sequence 170453, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 170453  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(131)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_124935C.1.pep  
 US-10-424-599-170453  
 Query Match 89.7%; Score 35; DB 12; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 EHDGIN 7  
 Db 106 EHDGIN 111  
 RESULT 8  
 US-10-424-599-284592  
 ; Sequence 284592, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 284592  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99011C.1.pep  
US-10-424-599-284592

Query Match 84.6%; Score 33; DB 12; Length 76;  
Best Local Similarity 83.3%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
Db 54 EHDGLN 59

RESULT 9  
US-10-083-357-1105  
Sequence 1105, Application US/10083357  
Publication No. US20030054370A1  
GENERAL INFORMATION:  
APPLICANT: Qiaodong Zeng et al.  
TITLE OF INVENTION: Systemic Discovery of New Genes  
FILE REFERENCE: 032796-090  
CURRENT APPLICATION NUMBER: US/10/083,357  
CURRENT FILING DATE: 2002-02-27  
NUMBER OF SEQ ID NOS: 1346  
SEQ ID NO 1105  
LENGTH: 17  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1105

Query Match 82.1%; Score 32; DB 14; Length 17;  
Best Local Similarity 71.4%; Pred. No. 9.5;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
Db 5 LSHDGLN 11

RESULT 10  
US-09-270-983-4  
Sequence 4, Application US/09270983  
Publication No. US20020132327A1  
GENERAL INFORMATION:  
APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY  
APPLICANT: HAY, Bruce A.  
TITLE OF INVENTION: METHOD FOR IDENTIFYING PROTEASES, PROTEASE TARGET SITES AND REGULATING PROTEASE ACTIVITY IN LIVING CELLS  
FILE REFERENCE: CIT1130-1  
CURRENT APPLICATION NUMBER: US/09/270,983  
CURRENT FILING DATE: 1999-03-17  
PRIOR APPLICATION NUMBER: US 60/078,721  
PRIOR FILING DATE: 1998-03-20  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 30  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Product Synthesis, such as solid phase synthesis  
US-09-270-983-4

Query Match 79.5%; Score 31; DB 12; Length 30;  
Best Local Similarity 83.3%; Pred. No. 28;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 284592  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_5334C.1.pep  
US-10-424-599-234026

Query Match 79.5%; Score 31; DB 12; Length 66;  
Best Local Similarity 71.4%; Pred. No. 66;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGI 6  
Db 11 IEHDGI 16

RESULT 11  
US-10-424-599-234026  
Sequence 234026, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 234026  
LENGTH: 66  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_5334C.1.pep  
US-10-424-599-234026

Query Match 79.5%; Score 31; DB 12; Length 66;  
Best Local Similarity 71.4%; Pred. No. 66;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
Db 23 LEHQGLN 29

RESULT 12  
US-09-747-287-203  
Sequence 203, Application US/09747287  
Publication No. US20030207264A1  
GENERAL INFORMATION:  
APPLICANT: KOMORIYA, AKIRA  
APPLICANT: PACKARD, BEVERLY S.  
TITLE OF INVENTION: HOMO-DOUBLY LABELED COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL SAMPLES  
FILE REFERENCE: 300-948600US  
CURRENT APPLICATION NUMBER: US/09/747,287  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/349,019  
PRIOR FILING DATE: 1999-09-10  
PRIOR APPLICATION NUMBER: US08/802,981  
PRIOR FILING DATE: 1997-02-20  
PRIOR APPLICATION NUMBER: PCT/US00/24882  
PRIOR FILING DATE: 2000-09-11  
NUMBER OF SEQ ID NOS: 242  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 203  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Artificial  
FEATURE:  
OTHER INFORMATION: Synthetic peptide.  
NAME/KEY: misc.feature  
LOCATION: (4)..(4)  
OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
NAME/KEY: misc.feature  
LOCATION: (14)..(14)  
OTHER INFORMATION: Xaa is epsilon-aminocaproic acid  
US-09-747-287-203

Query Match 76.9%; Score 30; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
 |||||  
 Db 8 HDGIN 12

RESULT 13  
 US-09-874-350A-169  
 ; Sequence 169, Application US/09874350A  
 ; Publication No. US2004009626A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oncoimmun, Inc.  
 ; APPLICANT: Komoriya, Akira  
 ; APPLICANT: Packard, Beverly  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL S  
 ; FILE REFERENCE: 300-903840US  
 ; CURRENT APPLICATION NUMBER: US/09/874,350A  
 ; CURRENT FILING DATE: 2001-05-04  
 ; PRIOR APPLICATION NUMBER: PCT/US98/00300  
 ; PRIOR FILING DATE: 1998-02-20  
 ; PRIOR APPLICATION NUMBER: PCT/US00/24882  
 ; PRIOR FILING DATE: 2000-09-11  
 ; PRIOR APPLICATION NUMBER: US 09/394,019  
 ; PRIOR FILING DATE: 1999-09-10  
 ; PRIOR APPLICATION NUMBER: US 08/802,981  
 ; PRIOR FILING DATE: 1997-02-20  
 ; NUMBER OF SEQ ID NOS: 221  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 169  
 ; LENGTH: 18  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic peptide substrate  
 ; NAME/KEY: MOD.RES  
 ; LOCATION: (1)..(1)  
 ; OTHER INFORMATION: K is blocked with Fmoc  
 ; NAME/KEY: MOD.RES  
 ; LOCATION: (14)..(14)  
 ; OTHER INFORMATION: X is epsilon-aminocaproic acid  
 ; NAME/KEY: MOD.RES  
 ; LOCATION: (4)..(4)  
 ; OTHER INFORMATION: X is epsilon aminocaproic acid  
 ; US-09-874-350A-169

Query Match 76.9%; Score 30; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
 |||||  
 Db 8 HDGIN 12

RESULT 14  
 US-10-424-599-210542  
 ; Sequence 210542, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 210542  
 ; LENGTH: 55  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_32146C.1.pep  
 ; US-10-424-599-210542

Query Match 76.9%; Score 30; DB 12; Length 55;  
 Best Local Similarity 57.1%; Pred. No. 86;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
 :|||:  
 Db 19 ISHDGIN 25

RESULT 15  
 US-10-424-599-160368  
 ; Sequence 160368, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 160368  
 ; LENGTH: 70  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(70)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_115630C.1.pep  
 ; US-10-424-599-160368

Query Match 76.9%; Score 30; DB 12; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 11e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
 |||||  
 Db 19 HDGIN 23

RESULT 16  
 US-10-424-599-221259  
 ; Sequence 221259, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 221259  
 ; LENGTH: 121  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:

NAME/KEY: unsure  
LOCATION: (1)..(121)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_41827C.1.pep  
US-10-424-599-221259

Query Match 76.9%; Score 30; DB 12; Length 121;  
Best Local Similarity 57.1%; Pred. No. 2e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDCIN 7  
Db 77 MSHDGLN 83

RESULT 17  
US-10-425-114-70959  
Sequence 70959, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 70959  
LENGTH: 128  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-ZMFLMOL7238F01\_FLI.pep  
US-10-425-114-70959

Query Match 76.9%; Score 30; DB 12; Length 128;  
Best Local Similarity 71.4%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDCIN 7  
Db 16 LHHEGIN 22

RESULT 18  
US-09-947-387-73  
Sequence 73, Application US/09947387  
Patent No. US20020150885A1  
GENERAL INFORMATION:  
APPLICANT: Weber, Eckard  
APPLICANT: Cai, Sui Xiong  
APPLICANT: Keana, John F.W.  
APPLICANT: Drewe, John A.  
APPLICANT: Zhang, Han-Zhong  
TITLE OF INVENTION: No. US20020150885A1el Fluorogenic or Fluorescent Reporter Molecule  
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence  
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the  
TITLE OF INVENTION: Use Thereof  
FILE REFERENCE: 1735.0290005  
CURRENT APPLICATION NUMBER: US/09/947,387  
CURRENT FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: US 60/061,582  
PRIOR FILING DATE: 1997-10-10  
PRIOR APPLICATION NUMBER: US 60/145,746  
PRIOR FILING DATE: 1998-03-03  
PRIOR APPLICATION NUMBER: US 09/168,888  
PRIOR FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 142  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 73  
LENGTH: 5  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Peptide  
US-09-947-387-73

Query Match 74.4%; Score 29; DB 9; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDCG 5  
Db 1 LEHDCG 5

RESULT 19  
US-09-747-287-200  
Sequence 200, Application US/09747287  
Publication No. US20030207264A1  
GENERAL INFORMATION:  
APPLICANT: KOMORIYA, AKIRA  
APPLICANT: PACKARD, BEVERLY S.  
TITLE OF INVENTION: HOMO-DOUBLY LABELED COMPOSITIONS FOR THE DETECTION OF ENZYME  
TITLE OF INVENTION: ACTIVITY IN BIOLOGICAL SAMPLES  
FILE REFERENCE: 300-948600US  
CURRENT APPLICATION NUMBER: US/09/747,287  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/349,019  
PRIOR FILING DATE: 1999-09-10  
PRIOR APPLICATION NUMBER: US08/802,981  
PRIOR FILING DATE: 1997-02-20  
PRIOR APPLICATION NUMBER: PCT/US00/24882  
PRIOR FILING DATE: 2000-09-11  
NUMBER OF SEQ ID NOS: 242  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 200  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Artificial  
FEATURE:  
OTHER INFORMATION: Synthetic peptide.  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (4)..(4)  
OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (14)..(14)  
OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
US-09-747-287-200

Query Match 74.4%; Score 29; DB 12; Length 18;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDCIN 7  
Db 6 LETDCIN 12

RESULT 20  
US-09-874-350A-166  
Sequence 166, Application US/09874350A  
Publication No. US2004009626A1  
GENERAL INFORMATION:  
APPLICANT: Oncoimmunin, Inc.  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly

;; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL S

;; TITLE OF INVENTION: COMPOSITIONS FOR THE DETECTION OF ENZYME ACTIVITY IN BIOLOGICAL S

;; FILE REFERENCE: 300-903840US  
;; CURRENT APPLICATION NUMBER: US/09/874,350A  
;; PRIOR FILING DATE: 2001-06-04  
;; PRIOR APPLICATION NUMBER: PCT/US98/00300  
;; PRIOR FILING DATE: 1998-02-20  
;; PRIOR APPLICATION NUMBER: PCT/US00/24882  
;; PRIOR FILING DATE: 2000-09-11  
;; PRIOR APPLICATION NUMBER: US 09/394,019  
;; PRIOR FILING DATE: 1999-09-10  
;; PRIOR APPLICATION NUMBER: US 08/802,981  
;; PRIOR FILING DATE: 1997-02-20  
;; NUMBER OF SEQ ID NOS: 221  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 166  
;; LENGTH: 18  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Synthetic peptide substrate  
;; NAME/KEY: MOD RES  
;; LOCATION: (1)..(1)  
;; OTHER INFORMATION: K is blocked with Fmoc  
;; NAME/KEY: MOD RES  
;; LOCATION: (14)..(14)  
;; OTHER INFORMATION: X is epsilon-aminocaproic acid  
;; NAME/KEY: MOD RES  
;; LOCATION: (4)..(4)  
;; OTHER INFORMATION: X is epsilon aminocaproic acid  
US-09-874-350A-166

;; FILE REFERENCE: 300-903840US  
;; CURRENT APPLICATION NUMBER: US/10294443  
;; PUBLICATION NO. US20040023865A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Reed, Steven G.  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Houghton, Raymond L.  
;; APPLICANT: Sleath, Paul R.  
;; APPLICANT: McNeill, Patricia D.  
;; APPLICANT: Homer, Mary J.  
;; APPLICANT: Secrist, Heather  
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
;; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
;; FILE REFERENCE: 210121.426C12  
;; CURRENT APPLICATION NUMBER: US/10/294,443  
;; CURRENT FILING DATE: 2002-11-13  
;; NUMBER OF SEQ ID NOS: 243  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 215  
;; LENGTH: 20  
;; TYPE: PRT  
;; ORGANISM: Babesia microti  
US-10-294-443-215

Query Match 74.4%; Score 29; DB 12; Length 18;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query Match 74.4%; Score 29; DB 16; Length 20;  
Best Local Similarity 83.3%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
Db 6 LETDGIN 12

QY 2 EHDGIN 7  
Db 6 EHDNIN 11

RESULT 21  
US-09-853-079-215  
;; Sequence 215, Application US/09853079  
;; Publication No. US20030109689A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Reed, Steven G.  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Houghton, Raymond L.  
;; APPLICANT: Sleath, Paul R.  
;; APPLICANT: McNeill, Patricia D.  
;; APPLICANT: Homer, Mary  
;; APPLICANT: Secrist, Heather  
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
;; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
;; FILE REFERENCE: 210121.426C11  
;; CURRENT APPLICATION NUMBER: US/09/853,079  
;; CURRENT FILING DATE: 2001-05-09  
;; NUMBER OF SEQ ID NOS: 224  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 215  
;; LENGTH: 20  
;; TYPE: PRT  
;; ORGANISM: Babesia microti  
US-09-853-079-215

US-09-867-550-774  
;; Sequence 774, Application US/09867550  
;; Patent No. US20020082206A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Leach, Martin D.  
;; APPLICANT: Mehraban, Fuad,  
;; APPLICANT: Conley, Pamela  
;; APPLICANT: Law, Debbie  
;; APPLICANT: Topper, James  
;; TITLE OF INVENTION: Thereby  
;; FILE REFERENCE: 21402-013 (Cura-313)  
;; CURRENT APPLICATION NUMBER: US/09/867,550  
;; CURRENT FILING DATE: 2001-09-20  
;; PRIOR APPLICATION NUMBER: USSN 60/208,427  
;; PRIOR FILING DATE: 2000-05-30  
;; NUMBER OF SEQ ID NOS: 2125  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 774  
;; LENGTH: 84  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-867-550-774

Query Match 74.4%; Score 29; DB 10; Length 20;  
Best Local Similarity 83.3%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query Match 74.4%; Score 29; DB 9; Length 84;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
Db 6 EHDNIN 11

QY 1 LEHDG 5  
Db 29 LEHDG 33

RESULT 24  
US-10-424-599-165583  
;; Sequence 165583, Application US/10424599

US-10-294-443-215  
;; Sequence 215, Application US/10294443  
;; Publication No. US20040023865A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Reed, Steven G.  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Houghton, Raymond L.  
;; APPLICANT: Sleath, Paul R.  
;; APPLICANT: McNeill, Patricia D.  
;; APPLICANT: Homer, Mary J.  
;; APPLICANT: Secrist, Heather  
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
;; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
;; FILE REFERENCE: 210121.426C12  
;; CURRENT APPLICATION NUMBER: US/10/294,443  
;; CURRENT FILING DATE: 2002-11-13  
;; NUMBER OF SEQ ID NOS: 243  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 215  
;; LENGTH: 20  
;; TYPE: PRT  
;; ORGANISM: Babesia microti  
US-10-294-443-215

Publication No. US20040031072A1  
GENERAL INFORMATION: Thomas J  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 165583  
LENGTH: 98  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(98)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_120537C.1.pep  
US-10-424-599-165583

Query Match 74.4%; Score 29; DB 12; Length 98;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|||||  
Db 80 LEHDG 84

RESULT 25  
US-10-424-599-203990  
Sequence 203990, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION: Thomas J  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 203990  
LENGTH: 101  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_2622C.1.pep  
US-10-424-599-203990

Query Match 74.4%; Score 29; DB 12; Length 101;  
Best Local Similarity 66.7%; Pred. No. 2.7e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
|||||  
Db 55 EHDGVD 60

RESULT 26  
US-10-369-493-11679  
Sequence 11679, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.

Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 165583  
LENGTH: 98  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(98)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_120537C.1.pep  
US-10-424-599-165583

Query Match 74.4%; Score 29; DB 12; Length 98;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|||||  
Db 80 LEHDG 84

RESULT 25  
US-10-424-599-203990  
Sequence 203990, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION: Thomas J  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 203990  
LENGTH: 101  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_2622C.1.pep  
US-10-424-599-203990

Query Match 74.4%; Score 29; DB 12; Length 101;  
Best Local Similarity 66.7%; Pred. No. 2.7e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
|||||  
Db 55 EHDGVD 60

RESULT 26  
US-10-369-493-11679  
Sequence 11679, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.

Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 11679  
LENGTH: 127  
TYPE: PRT  
ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-11679

Query Match 74.4%; Score 29; DB 15; Length 127;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|||||  
Db 93 LEHDG 97

RESULT 27  
US-10-369-493-14258  
Sequence 14258, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 14258  
LENGTH: 127  
TYPE: PRT  
ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-14258

Query Match 74.4%; Score 29; DB 15; Length 127;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|||||  
Db 93 LEHDG 97

RESULT 28  
US-10-369-493-14694  
Sequence 14694, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28



; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 14694  
; LENGTH: 127  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-14694

Query Match 74.4%; Score 29; DB 15; Length 127;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
| | | | |  
Db 93 LEHDG 97

RESULT 29  
US-10-424-599-147562  
; Sequence 147562, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, PAUL  
; APPLICANT: GENMELL, JACK  
; APPLICANT: AVERBACK, PAUL  
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
; FILE REFERENCE: 59003.000008  
; CURRENT APPLICATION NUMBER: US/10/198,070  
; CURRENT FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 60/306,161  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/306,150  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/331,477  
; PRIOR FILING DATE: 2001-11-16

Query Match 74.4%; Score 29; DB 15; Length 127;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
| | | | |  
Db 93 LEHDG 97

RESULT 29  
US-10-424-599-147562  
; Sequence 147562, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, PAUL  
; APPLICANT: GENMELL, JACK  
; APPLICANT: AVERBACK, PAUL  
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
; FILE REFERENCE: 59003.000008  
; CURRENT APPLICATION NUMBER: US/10/198,070  
; CURRENT FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 60/306,161  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/306,150  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/331,477  
; PRIOR FILING DATE: 2001-11-16

Query Match 74.4%; Score 29; DB 15; Length 127;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
| | | | |  
Db 93 LEHDG 97

RESULT 29  
US-10-424-599-147562  
; Sequence 147562, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, PAUL  
; APPLICANT: GENMELL, JACK  
; APPLICANT: AVERBACK, PAUL  
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
; FILE REFERENCE: 59003.000008  
; CURRENT APPLICATION NUMBER: US/10/198,070  
; CURRENT FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 60/306,161  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/306,150  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/331,477  
; PRIOR FILING DATE: 2001-11-16

; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-198-070-26

Query Match 74.4%; Score 29; DB 14; Length 137;  
Best Local Similarity 57.1%; Pred. No. 3.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
| | | | |  
Db 109 LQHQGVN 115

RESULT 31  
US-10-422-571-19  
; Sequence 19, Application US/10422571  
; Publication No. US20040083497A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Paris, Mary  
; APPLICANT: Ge, Wangmao  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20082.00  
; CURRENT APPLICATION NUMBER: US/10/422,571  
; CURRENT FILING DATE: 2003-04-23  
; PRIOR APPLICATION NUMBER: US60/404,306  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US60/423,290  
; PRIOR FILING DATE: 2002-11-01  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-422-571-19

Query Match 74.4%; Score 29; DB 16; Length 137;  
Best Local Similarity 57.1%; Pred. No. 3.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
| | | | |  
Db 109 LQHQGVN 115

RESULT 32  
US-10-422-571-29  
; Sequence 29, Application US/10422571  
; Publication No. US20040083497A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Paris, Mary  
; APPLICANT: Ge, Wangmao  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20082.00  
; CURRENT APPLICATION NUMBER: US/10/422,571  
; CURRENT FILING DATE: 2003-04-23  
; PRIOR APPLICATION NUMBER: US60/404,306

```

; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US60/423,290
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-422-571-29

Query Match      74.4%; Score 29; DB 16; Length 137;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDCIN 7
Db 109 LQHQGVN 115

RESULT 33
US-10-422-571-38
; Sequence 38, Application US/10422571
; Publication No. US20040083497A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Faris, Mary
; APPLICANT: Ge, Wangmao
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20082.00
; CURRENT FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US60/404,306
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US60/423,290
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-422-571-38

Query Match      74.4%; Score 29; DB 16; Length 137;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDCIN 7
Db 109 LQHQGVN 115

RESULT 34
US-10-422-571-39
; Sequence 39, Application US/10422571
; Publication No. US20040083497A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Faris, Mary
; APPLICANT: Ge, Wangmao
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20082.00

```

---

```

; CURRENT APPLICATION NUMBER: US/10/422,571
; CURRENT FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US60/404,306
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US60/423,290
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-422-571-39

Query Match      74.4%; Score 29; DB 16; Length 137;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDCIN 7
Db 109 LQHQGVN 115

RESULT 35
US-10-422-571-89
; Sequence 89, Application US/10422571
; Publication No. US20040083497A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Faris, Mary
; APPLICANT: Ge, Wangmao
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins
; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20082.00
; CURRENT APPLICATION NUMBER: US/10/422,571
; CURRENT FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: US60/404,306
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US60/423,290
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-422-571-89

Query Match      74.4%; Score 29; DB 16; Length 137;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDCIN 7
Db 109 LQHQGVN 115

RESULT 36
US-10-422-571-126
; Sequence 126, Application US/10422571
; Publication No. US20040083497A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Jakobovits, Aya
; APPLICANT: Faris, Mary
; APPLICANT: Ge, Wangmao
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins

```

; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20082.00  
; CURRENT APPLICATION NUMBER: US/10/422,571  
; CURRENT FILING DATE: 2003-04-23  
; PRIOR APPLICATION NUMBER: US60/404,306  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US60/423,290  
; PRIOR FILING DATE: 2002-11-01  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 126  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-422-571-126

Query Match 74.4%; Score 29; DB 16; Length 137;  
Best Local Similarity 57.1%; Pred. No. 3.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHGDGIN 7  
|:|:|:  
Db 109 LQHQQGVN 115

RESULT 37  
US-10-422-571-130  
; Sequence 130, Application US/10422571  
; Publication No. US20040083497A1  
; GENERAL INFORMATION:  
; APPLICANT: Agensys, Inc.  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Challita-Eid, Pia M.  
; APPLICANT: Jakobovits, Aya  
; APPLICANT: Paris, Mary  
; APPLICANT: Ge, Wangmao  
; TITLE OF INVENTION: Nucleic Acids and Corresponding Proteins  
; TITLE OF INVENTION: Entitled 191P4D12(b) Useful in Treatment and Detection of  
; FILE REFERENCE: 51158-20082.00  
; CURRENT APPLICATION NUMBER: US/10/422,571  
; CURRENT FILING DATE: 2003-04-23  
; PRIOR APPLICATION NUMBER: US60/404,306  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US60/423,290  
; PRIOR FILING DATE: 2002-11-01  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 130  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-422-571-130

Query Match 74.4%; Score 29; DB 16; Length 137;  
Best Local Similarity 57.1%; Pred. No. 3.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHGDGIN 7  
|:|:|:  
Db 109 LQHQQGVN 115

RESULT 38  
US-10-221-278-647  
; Sequence 647, Application US/10221278  
; Publication No. US20040034208A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-045  
; CURRENT APPLICATION NUMBER: US/10/221,278

; CURRENT FILING DATE: 2002-09-06  
; PRIOR APPLICATION NUMBER: 09/693,267  
; PRIOR FILING DATE: 2000-10-20  
; CURRENT APPLICATION NUMBER: 09/665,363  
; PRIOR FILING DATE: 2000-09-19  
; PRIOR APPLICATION NUMBER: 09/616,847  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 09/596,193  
; PRIOR FILING DATE: 2000-06-17  
; PRIOR APPLICATION NUMBER: 09/574,454  
; PRIOR FILING DATE: 2000-05-19  
; PRIOR APPLICATION NUMBER: 09/519,705  
; PRIOR FILING DATE: 2000-03-07  
; NUMBER OF SEQ ID NOS: 752  
; SEQ ID NO 647  
; LENGTH: 143  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-221-278-647

Query Match 74.4%; Score 29; DB 12; Length 143;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGI 6  
|:|:|:  
Db 63 EHDGI 67

RESULT 39  
US-10-291-172-647  
; Sequence 647, Application US/10291172  
; Publication No. US20030228584A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-045  
; CURRENT APPLICATION NUMBER: US/10/291,172  
; CURRENT FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 09/693,267  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 09/665,363  
; PRIOR FILING DATE: 2000-09-19  
; PRIOR APPLICATION NUMBER: 09/616,847  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 09/596,193  
; PRIOR FILING DATE: 2000-06-17  
; PRIOR APPLICATION NUMBER: 09/574,454  
; PRIOR FILING DATE: 2000-05-19  
; PRIOR APPLICATION NUMBER: 09/519,705  
; PRIOR FILING DATE: 2000-03-07  
; NUMBER OF SEQ ID NOS: 752  
; SEQ ID NO 647  
; LENGTH: 143  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-291-172-647

Query Match 74.4%; Score 29; DB 15; Length 143;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHDGI 6  
|:|:|:  
Db 63 EHDGI 67

RESULT 40  
US-10-424-599-262285  
; Sequence 262285, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J

```
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 262285
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78865C.1.pep
US-10-424-599-262285

Query Match          71.8%; Score 28; DB 12; Length 49;
Best Local Similarity 56.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EHDGIN 7
    |||
Db 28 EHGVN 33

RESULT 41
US-09-864-408A-116
; Sequence 116, Application US/09864408A
; Publication No. US2004000947A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US2004000947A1e1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 116
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-116

Query Match          71.8%; Score 28; DB 11; Length 52;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7
    |||
Db 32 HDGLN 36

RESULT 42
US-10-424-599-208279
; Sequence 208279, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 208279
; LENGTH: 55

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 208279
; LENGTH: 55

; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_36300C.1.pep
US-10-424-599-208279

Query Match          71.8%; Score 28; DB 12; Length 55;
Best Local Similarity 57.1%; Pred. No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7
    :|||
Db 46 MQHGIN 52

RESULT 43
US-10-424-599-222247
; Sequence 222247, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 222247
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_42718C.1.pep
US-10-424-599-222247

Query Match          71.8%; Score 28; DB 12; Length 55;
Best Local Similarity 65.7%; Pred. No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDGI 6
    :|||
Db 38 LQHDGL 43

RESULT 44
US-10-424-599-215145
; Sequence 215145, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 215145
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_36300C.1.pep
US-10-424-599-215145

Query Match          71.8%; Score 28; DB 12; Length 68;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 LEHDCI 6  
Db 34 LSHDCI 39

## RESULT 45

US-10-424-599-216092  
; Sequence 216092, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 216092  
; LENGTH: 70  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_37160C.1.pep  
US-10-424-599-216092

Query Match 71.8%; Score 28; DB 12; Length 70;  
Best Local Similarity 66.7%; Fred. No. 2.9e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
Db 35 EHDGLD 40

Search completed: June 15, 2004, 12:27:25  
Job time : 43 secs



R.Yayon, A.; Zimmer, Y.; Guo-Hong, S.; Avivi, A.; Yarden, Y.; Givol, D.  
EMBO J. 11, 1885-1890, 1992  
A:Title: A confined variable region confers ligand specificity on fibroblast growth factor  
A:Reference number: S20961; MUID:92258398; PMID:1316275  
A:Accession: S20961  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-64 <YAY>  
C:Keywords: growth factor receptor

Query Match 74.4%; Score 29; DB 2; Length 64;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:|:|  
Db 1 LEHDSGIN 7

RESULT 6  
F90337  
hypothetical protein SSO1764 [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: F90337  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arratt, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: F90337  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3815007; PIDN:AAK41957.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO1764

Query Match 74.4%; Score 29; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDSG 5  
|:|:|:|  
Db 101 LEHDSG 105

RESULT 7  
B46078  
POU protein Emb, form c7 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 15-Oct-1999  
C:Accession: B46078  
R:Okamoto, K.; Wakamiya, M.; Noji, S.; Koyama, E.; Taniguchi, S.; Takemura, R.; Copeland,  
J. Biol. Chem. 269, 7449-7457, 1993  
A:Title: A novel class of murine POU gene predominantly expressed in central nervous system  
A:Reference number: A46078; MUID:93216691; PMID:8463278  
A:Accession: B46078  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-44 <OKA>  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 71.8%; Score 28; DB 2; Length 44;  
Best Local Similarity 71.4%; Pred. No. 32;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:|:|  
Db 26 LEHDSGIN 32

R.Yayon, A.; Zimmer, Y.; Guo-Hong, S.; Avivi, A.; Yarden, Y.; Givol, D.  
EMBO J. 11, 1885-1890, 1992  
A:Title: A confined variable region confers ligand specificity on fibroblast growth factor  
A:Reference number: S20961; MUID:92258398; PMID:1316275  
A:Accession: S20961  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-64 <YAY>  
C:Keywords: growth factor receptor

Query Match 74.4%; Score 29; DB 2; Length 64;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:|:|  
Db 1 LEHDSGIN 7

RESULT 6  
F90337  
hypothetical protein SSO1764 [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: F90337  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arratt, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: F90337  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3815007; PIDN:AAK41957.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO1764

Query Match 74.4%; Score 29; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDSG 5  
|:|:|:|  
Db 101 LEHDSG 105

RESULT 7  
B46078  
POU protein Emb, form c7 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 15-Oct-1999  
C:Accession: B46078  
R:Okamoto, K.; Wakamiya, M.; Noji, S.; Koyama, E.; Taniguchi, S.; Takemura, R.; Copeland,  
J. Biol. Chem. 269, 7449-7457, 1993  
A:Title: A novel class of murine POU gene predominantly expressed in central nervous system  
A:Reference number: A46078; MUID:93216691; PMID:8463278  
A:Accession: B46078  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-44 <OKA>  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 71.8%; Score 28; DB 2; Length 44;  
Best Local Similarity 71.4%; Pred. No. 32;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:|:|  
Db 26 LEHDSGIN 32

R.Yayon, A.; Zimmer, Y.; Guo-Hong, S.; Avivi, A.; Yarden, Y.; Givol, D.  
EMBO J. 11, 1885-1890, 1992  
A:Title: A confined variable region confers ligand specificity on fibroblast growth factor  
A:Reference number: S20961; MUID:92258398; PMID:1316275  
A:Accession: S20961  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-64 <YAY>  
C:Keywords: growth factor receptor

Query Match 74.4%; Score 29; DB 2; Length 64;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:|:|  
Db 1 LEHDSGIN 7

RESULT 6  
F90337  
hypothetical protein SSO1764 [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: F90337  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arratt, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: F90337  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3815007; PIDN:AAK41957.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO1764

Query Match 74.4%; Score 29; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDSG 5  
|:|:~|:|  
Db 101 LEHDSG 105

RESULT 7  
B46078  
POU protein Emb, form c7 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 15-Oct-1999  
C:Accession: B46078  
R:Okamoto, K.; Wakamiya, M.; Noji, S.; Koyama, E.; Taniguchi, S.; Takemura, R.; Copeland,  
J. Biol. Chem. 269, 7449-7457, 1993  
A:Title: A novel class of murine POU gene predominantly expressed in central nervous system  
A:Reference number: A46078; MUID:93216691; PMID:8463278  
A:Accession: B46078  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-44 <OKA>  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 71.8%; Score 28; DB 2; Length 44;  
Best Local Similarity 71.4%; Pred. No. 32;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:~|:|  
Db 26 LEHDSGIN 32

R.Yayon, A.; Zimmer, Y.; Guo-Hong, S.; Avivi, A.; Yarden, Y.; Givol, D.  
EMBO J. 11, 1885-1890, 1992  
A:Title: A confined variable region confers ligand specificity on fibroblast growth factor  
A:Reference number: S20961; MUID:92258398; PMID:1316275  
A:Accession: S20961  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-64 <YAY>  
C:Keywords: growth factor receptor

Query Match 74.4%; Score 29; DB 2; Length 64;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDSGIN 7  
|:|:~|:|  
Db 1 LEHDSGIN 7

RESULT 6  
F90337  
hypothetical protein SSO1764 [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: F90337  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arratt, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.

```

Query Match          71.8%;      Score 28;      DB 2;      Length 118;
Best Local Similarity 56.7%;      Pred. No. 1e-02;
Matches 4;      Conservative 2;      Mismatches 0;      Gaps 0;

QY      1 LEHDMI 6
      |||:::
      |||:::
DB      18 LEHEGV 23

RESULT 10
ERAD42
early E3B 14K protein - human adenovirus 2
C:Species: Mastadenovirus h2 (human adenovirus 2)
A:Note: Host Homo sapiens (man)
C:Date: 29-Jul-1981 #sequence_revision 08-Oct-1981 #text_change 04-Mar-1994
C:Accession: A03818
R:Herisse, J.; Galibert, F.
Nucleic Acids Res. 9, 1229-1240, 1981
A:Title: Nucleotide sequence of the EcoRI E fragment of adenovirus 2 genome.
A:Reference number: A93722; MUID:81198955; PMID:6282722
A:Accession: A03818
A:Molecule type: DNA
A:Residues: 1-128 <HER>

```

**RESULT 12**

T12911  
hypothetical protein yorZ - Bacillus subtilis phage SPBc2  
C/Species: Bacillus subtilis phage SPBc2  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
C/Accession: T12911; B69925  
R/Lazarovic, V.; Duesterhoef, A.; Solido, B.; Hilbert, H.; MaueI, C.; Karamata, D.  
submitted to the EMBL Data Library, August 1997  
A/Description: The complete nucleotide sequence of the Bacillus subtilis SPBetac2 prophage  
A/Reference number: Z17583  
A/Accession: T12911  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-70 <LAZ>  
A/Cross-references: EMBL:AF020713; NID:G3025478; PID:G3025625; PIDN:RAC13120..1  
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Bruschii, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choj  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallert  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,



A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yamoto, K.; Yata, K.; Yoshida, K  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: B69925  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-70 <KUN>  
A;Cross-references: GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13912.1; PID:e1185492;  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: yorZ

Query Match 69.2%; Score 27; DB 2; Length 70;  
Best Local Similarity 71.4%; Pred. No. 90;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
|||  
Db 57 LEHGLN 63

RESULT 13  
T29085  
hypothetical protein SC12.36 - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C;Accession: T29085  
R;Redenbach, M.; Kiese, H.M.; Denapate, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw  
Mol. Microbiol. 21, 77-96, 1996  
A;Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S  
A;Reference number: Z20556; MUID:97000351; PMID:8843436  
A;Accession: T29085  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-76 <RED>  
A;Cross-references: EMBL:AL031124; NID:e1312893; PID:e1312923; PIDN:CAA20003.1  
C;Genetics:  
A;Note: SC12.36

Query Match 69.2%; Score 27; DB 2; Length 76;  
Best Local Similarity 80.0%; Pred. No. 99;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|||  
Db 1 MEHDG 5

RESULT 14  
B83798  
hypothetical protein BH1186 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: B83798  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: B83798  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-114 <STO>  
A;Cross-references: GB:AF001511; GB:BA000004; NID:g10173727; PIDN:BA04905.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH1186

Query Match 69.2%; Score 27; DB 2; Length 114;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDOI 6  
|||  
Db 50 MKHDGI 55

RESULT 15  
T20983  
hypothetical protein F15D4.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T20983  
R;Smye, R.  
submitted to the EMBL Data Library, September 1996  
A;Reference number: Z19354  
A;Accession: T20983  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-115 <WIL>  
A;Cross-references: EMBL:Z80344; PIDN:CAB02489.1; GSPDB:GN00020; CESP:F15D4.6  
A;Experimental source: clone F15D4  
C;Genetics:  
A;Gene: CESP:F15D4.6  
A;Map position: 2  
A;Introns: 41/2; 82/1

Query Match 69.2%; Score 27; DB 2; Length 115;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
|||  
Db 63 LENDGYN 69

RESULT 16  
F95121  
conserved domain protein SP1053 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C;Accession: F95121  
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: F95121  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-125 <KUR>  
A;Cross-references: GB:AE005672; PIDN:AAK75167.1; PID:g14972527; GSPDB:GN00164; TIGR:SP4  
A;Experimental source: strain TIGR4  
C;Genetics:  
A;Gene: SP1053

Query Match 69.2%; Score 27; DB 2; Length 125;  
Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
|||  
Db 77 MESDGVN 83

RESULT 17  
A48552  
protein US420 - infectious laryngotracheitis virus  
C;Species: infectious laryngotracheitis virus

C>Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 08-Oct-1999  
 C/Accession: A48552  
 R/Sakaguchi, M.; Urakawa, T.; Hirayama, Y.; Miki, N.; Yamamoto, M.; Hirai, K.  
 Virus Genes 6, 365-378, 1992  
 A>Title: Sequence determination and genetic content of an 8.9-kb restriction fragment in  
 A/Reference number: A48552; MUID:93118245; PMID:1282282  
 A/Accession: A48552  
 A>Status: Preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-140 <SAK>  
 A/Cross-references: GB:M80595; NID:G291557; PIDN:AAB59889.1; PID:G291558  
 A/Note: sequence extracted from NCBI backbone (NCBIN:121622, NCBIPI:121623)  
 C:Superfamily: infectious laryngotracheitis virus protein US420

Query Match 69.2%; Score 27; DB 2; Length 140;  
 Best Local Similarity 71.4%; Pred. No. 2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHGDGIN 7  
 |::|:  
 Db 35 LLEHGIN 41

RESULT 18  
 Ti6039  
 hypothetical protein F12A10.5 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Dec-1999  
 C/Accession: Ti6039  
 R/Geisel, C.  
 submitted to the EMBL Data Library, June 1995  
 A/Description: The sequence of C. elegans cosmid F12A10.  
 A/Reference number: Z18451  
 A/Accession: Ti6039  
 A>Status: Preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-145 <GEI>  
 A/Cross-references: EMBL:U28731; NID:G861241; PID:G861246; PIDN:AAA68298.1; CESP:F12A10.  
 A/Experimental source: strain Bristol N2  
 C/Genetics:  
 A/Gene: CESP:F12A10.5  
 A/Introns: 30/3; 56/1; 87/3; 137/1  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 C/Keywords: EF hand

Query Match 69.2%; Score 27; DB 2; Length 145;  
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHGDGI 6  
 |::|:  
 Db 127 LDHGDGM 132

RESULT 19  
 F97328  
 hypothetical protein CAC3491 [imported] - Clostridium acetobutylicum  
 C/Species: Clostridium acetobutylicum  
 C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C/Accession: F97328  
 R/Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A>Title: Genome sequence and comparative analysis of the Solvent-Producing Bacterium C10  
 A/Reference number: A96900; MUID:21359325; PMID:21359325  
 A/Accession: F97328  
 A>Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-145 <KUR>  
 A/Cross-references: GB:AE001437; PIDN:AAK81417.1; PID:G15026582; GSPDB:GN00168  
 A/Experimental source: Clostridium acetobutylicum ATCC824  
 C/Genetics:  
 A/Gene: CAC3491

Query Match 69.2%; Score 27; DB 2; Length 145;  
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDDGI 6  
 |::|:  
 Db 39 LKHDDGL 44

RESULT 20  
 B97991  
 hypothetical protein spr0954 [imported] - Streptococcus pneumoniae (strain R6)  
 C/Species: Streptococcus pneumoniae  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
 C/Accession: B97991  
 R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; Re,  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc,  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A/Reference number: A97872; MUID:21429245; PMID:11544234  
 A/Accession: B97991  
 A>Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-150 <KUR>  
 A/Cross-references: GB:AE007317; PIDN:AAK99758.1; PID:G15458564; GSPDB:GN00174  
 C/Genetics:  
 A/Gene: spr0954

Query Match 69.2%; Score 27; DB 2; Length 150;  
 Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDDGIN 7  
 |::|:  
 Db 102 MESDDGVN 108

RESULT 21  
 D81356  
 methylated-DNA-[protein]-cysteine S-methyltransferase (EC 2.1.1.63) Cj0836 [imported] - C  
 C/Species: Campylobacter jejuni  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C/Accession: D81356  
 R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling,  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell,  
 Nature 403, 665-668, 2000  
 A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
 A/Reference number: A81250; MUID:20150912; PMID:10698204  
 A/Accession: D81356  
 A>Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-150 <PAR>  
 A/Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAB73101.1; PID:G696828  
 A/Experimental source: serotype O2, strain NCIC 11168  
 C/Genetics:  
 A/Gene: OGT; Cj0836  
 C/Keywords: methyltransferase

Query Match 69.2%; Score 27; DB 2; Length 150;  
 Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LEHDDGIN 7  
 |::|:  
 Db 142 LENEDEVN 148

RESULT 22  
 A11911  
 30S ribosomal protein S21 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: A11911  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: A11911  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-58 <KUR>  
A:Cross-references: GB:BA000019; PID:BA072800.1; PID:g17130188; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: rps21  
C:Superfamily: Escherichia coli ribosomal protein S21

Query Match 66.7%; Score 26; DB 2; Length 58;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
|||:  
Db 8 EHEGID 13  
|||:  
|||:

RESULT 23  
H86911  
hypothetical protein [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: H86911  
R:Coile, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: H86911  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-83 <STO>  
A:Cross-references: GB:AL450380; NID:g13092434; PIDN:CAC29532.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: MLC024

Query Match 66.7%; Score 26; DB 2; Length 83;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
|||:  
Db 26 DHDGVD 31  
|||:  
|||:

RESULT 24  
AH2845  
Conserved hypothetical protein Atu2193 [imported] - Agrobacterium tumefaciens (strain C5  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AH2845  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
i Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AH2845

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-88 <KUR>  
A:Cross-references: GB:AB008688; PIDN:AAU43182.1; PID:g17740662; GSPDB:GN00186  
A:Experimental source: Strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu2193  
A:Map position: circular chromosome

Query Match 66.7%; Score 26; DB 2; Length 88;  
Best Local Similarity 80.0%; Pred. No. 1.9e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|||:  
Db 9 VEHDG 13  
|||:  
|||:

RESULT 25  
H82607  
conserved hypothetical protein XP2030 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 22-Jun-2003  
C:Accession: H82607  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: H82607  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-93 <SIM>  
A:Cross-references: GB:AB004021; GB:AB003849; NID:g9107139; PIDN:AAF84832.1; GSPDB:GN0011  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al  
Briones, M.E.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohne  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.B.; Kutamae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.N.; Miracca, E.C.; Miyaki, C.Y.  
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Teuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XP2030  
C:Superfamily: uncharacterized conserved protein sar2962

Query Match 66.7%; Score 26; DB 2; Length 93;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGI 6  
|||:  
Db 67 LEHDDI 72  
|||:  
|||:

RESULT 26  
AB2059  
hypothetical protein asl2024 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AB2059  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2059  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-97 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAF73723.1; PID:g17131115; GSPDB:GN00179  
A:Experimental source: strain POC 7120  
C:Genetics:  
A:Gene: asl2024

Query Match 66.7%; Score 26; DB 2; Length 97;  
Best Local Similarity 71.4%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDCIN 7  
||| :|||  
Db 36 LEKEGIN 42

RESULT 27  
I39650  
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - Actinobacillus pleuropneumoniae (fragments)  
C:Species: Actinobacillus pleuropneumoniae  
C:Date: 16-Aug-1996 #sequence revision 05-Mar-1999 #text\_change 20-Jun-2000  
A:Accession: I39650; S22815; S22161  
R:Kroll, J.S.; Langford, P.R.; Wilks, K.E.; Keil, A.D.  
Microbiology 141, 2271-2279, 1995  
A:Title: Bacterial [Cu,Zn]-superoxide dismutase: Phylogenetically distinct from the euka  
A:Reference number: I39485; MUID:96118708; PMID:7496539  
A:Accession: I39650  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-87 <KRO>  
A:Cross-references: EMBL:X83123; NID:g1019747; PIDN:CAA58204.1; PID:g1019748  
R:Loyns, B.M.; Langford, P.R.; Kroll, J.S.  
Nucleic Acids Res. 20, 615, 1992  
A:Title: recF in Actinobacillus pleuropneumoniae.  
A:Reference number: S22813; MUID:92158680; PMID:1741300  
A:Accession: S22815  
A:Molecule type: DNA  
A:Residues: 88-98 <LOY>  
A:Cross-references: EMBL:X63626; NID:g38951; PIDN:CAA45174.1; PID:g1333704  
A:Note: This sequence was submitted to the EMBL Data Library, December 1991  
A:Note: neither the complete nucleic acid sequence nor the complete translation are show  
C:Genetics:  
A:Gene: sodC  
C:Function:  
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
C:Superfamily: superoxide dismutase (Cu-Zn)  
C:Keywords: copper; metalloprotein; oxidoreductase; zinc  
F:18,27,36,39/Binding site: zinc (His, His, His, Asp) #status predicted  
F:91/Active site: Arg #status predicted

Query Match 66.7%; Score 26; DB 2; Length 98;  
Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDCG 5  
||| :|||  
Db 45 VEHDG 49

RESULT 28  
S78727  
protein YLL018c-a - Yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 15-Jan-1999 #sequence revision 15-Jan-1999 #text\_change 15-Jan-1999  
A:Accession: S78727  
R:Goffeau, A.; Fumelle, B.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64761  
A:Accession: S78727  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-98 <GOF>  
A:Cross-references: EMBL:Z73125; MIPS:YLL018c-a  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 12L

Query Match 66.7%; Score 26; DB 2; Length 98;  
Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
||| :|||  
Db 24 LDHDG 28

RESULT 29  
A31201  
GLI-related finger protein GLI2 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1990 #sequence revision 31-Mar-1990 #text\_change 31-Dec-1993  
A:Accession: A31201  
R:Ruppert, J.M.; Kinzler, K.W.; Wong, A.J.; Bigner, S.H.; Kao, F.T.; Law, M.L.; Seunanez, M.; Cell. Biol. 8, 3104-3113, 1988  
A:Title: The GLI-Kruppel family of human genes.  
A:Reference number: A93103; MUID:89096896; PMID:2850480  
A:Accession: A31201  
A:Molecule type: DNA  
A:Residues: 1-100 <RUP>  
C:Genetics:  
A:Introns: 50/2  
C:Superfamily: gli transforming protein  
C:Keywords: DNA binding; zinc finger

Query Match 66.7%; Score 26; DB 2; Length 100;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EHDGIN 7  
||| :|||  
Db 75 EHEGON 80

RESULT 30  
A97623  
hypothetical protein AGR\_C3987 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence revision 30-Sep-2001 #text\_change 18-Nov-2002  
A:Accession: A97623  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: A97623  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-101 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87938.1; PID:g15157340; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C3987  
A:Map position: circular chromosome

Query Match 66.7%; Score 26; DB 2; Length 101;  
Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
||| :|||  
Db 22 VEHDG 26

RESULT 31

H86483  
probable 60S ribosomal protein L30 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Nov-2001  
C;Accession: H86483  
R;Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: H86483  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-112 <STO>  
A;Cross-references: GB:AE005172; NID:gl0645486; PIDN:AAG21599.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: SMB20293  
A;Map position: 1  
C;Superfamily: rat ribosomal protein L30

Query Match 66.7%; Score 26; DB 2; Length 112;  
Best Local Similarity 80.0%; Pred. No. 2.5e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
|:|:|  
Db 11 HEGIN 15

RESULT 32  
AB2636  
conserved hypothetical protein Atu0485 [imported] - Agrobacterium tumefaciens (strain C58)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AB2636  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:111743193  
A;Accession: AB2636  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-117 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AAL41504.1; PID:g17738832; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: Atu0485  
A;Map position: circular chromosome  
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv2705c

Query Match 66.7%; Score 26; DB 2; Length 117;  
Best Local Similarity 80.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|:|:|  
Db 105 LDHDG 109

RESULT 33  
C95877  
hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSymB  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C;Accession: C95877  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: C95877  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-119 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC48683.1; PID:g15140155; GSPDB:GN00167  
A;Experimental source: strain 1021, megaplasmid pSymB  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pla, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
C;Genetics:  
A;Gene: SMB20293  
A;Genome: plasmid

Query Match 66.7%; Score 26; DB 2; Length 119;  
Best Local Similarity 80.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|:|:|  
Db 9 VEHDG 13

RESULT 34  
D87499  
hypothetical protein CC2017 [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C;Accession: D87499  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: D87499  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-130 <STO>  
A;Cross-references: GB:AE005673; NID:gl3423490; PIDN:AAK23992.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC2017

Query Match 66.7%; Score 26; DB 2; Length 130;  
Best Local Similarity 80.0%; Pred. No. 3e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|:|:|  
Db 68 LDHDG 72

RESULT 35  
T00002  
peps protein - Acidianus ambivalens plasmid pDL10  
C;Species: Acidianus ambivalens  
C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 20-Jun-2000  
C;Accession: T00002  
R;Kletzin, A.  
submitted to the EMBL Data Library, January 1998  
A;Description: The plasmid pDL10 from the extremely thermophilic archaeon Acidianus ambi  
A;Reference number: Z14035  
A;Accession: T00002

A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-135 <KLE>  
A:Cross-references: EMBL:AJ225333; PIDN:CAA12519.1  
A:Experimental source: strain Le110  
C:Genetics:  
A:Gene: pepG  
A:Genome: plasmid pDL10  
C:Superfamily: Acidianus ambivalens plasmid pDL10 pepG protein

Query Match 66.7%; Score 26; DB 2; Length 135;  
Best Local Similarity 71.4%; Pred. No. 3.1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7  
|||  
39 LERDGIS 45

Db 39 LERDGIS 45

RESULT 36  
C9276  
lek protein - Ectromelia virus  
C:Species: Ectromelia virus  
C:Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 08-Oct-1999  
A:Accession: C49276  
R:Senkevich, T.G.; Koonin, E.V.; Buller, R.M.L.  
Virology 198, 118-128, 1994  
A:Title: A poxvirus protein with a RING zinc finger motif is of crucial importance for  
A:Reference number: A49276; MUID:94082441; PMID:8259647  
A:Accession: C49276  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-138 <SEN>  
A:Cross-references: GB:U01161; NID:g397978; PIDN:AAA16259.1; PID:g397981  
C:Genetics:  
A:Gene: p16

Query Match 66.7%; Score 26; DB 2; Length 138;  
Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 1 LEH--DGIN 7  
|||  
68 IEHLGDGIN 75

Db 68 IEHLGDGIN 75

RESULT 37  
T17493  
citrate utilization determinant homolog A3R - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
A:Accession: T17493; T18191  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T17493  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-141 <GRA>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAAC96371.1  
A:Experimental source: specific host Chlorella strain NC64A  
A:Accession: T18191  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-141 <GR2>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAAC96990.1  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A:Gene: A3R  
A>Note: A689L

Query Match 66.7%; Score 26; DB 2; Length 141;  
Best Local Similarity 66.7%; Pred. No. 3.3e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 2; Gaps 0;

QY 1 LEHDMI 6  
|||  
85 LNHDL 90

Db 85 LNHDL 90

RESULT 38  
A54492  
hemoglobin beta-II chain, larval - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 04-Mar-2000  
A:Accession: A54492; A24119  
R:Meyerhof, W.; Koester, M.; Stalder, J.; Weber, R.; Knoechel, W.  
Mol Biol Rep. 11, 155-161, 1986  
A:Title: Sequence analysis of the larval beta-II-globin gene of Xenopus laevis.  
A:Reference number: A54492; MUID:87014375; PMID:3020391  
A:Accession: A54492  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-147 <MEY>  
A:Cross-references: GB:M21411; NID:g213961; PIDN:AAA49653.1; PID:g213962  
R:Knoechel, W.; Meyerhof, W.; Stalder, J.; Weber, R.  
Nucleic Acids Res. 13, 7899-7908, 1985  
A:Title: Comparative nucleotide sequence analysis of two types of larval beta-globin mRNA  
A:Reference number: A24119; MUID:86067205; PMID:2999708  
A:Accession: A24119  
A:Molecule type: mRNA  
A:Residues: 2-147 <KNO>  
A:Cross-references: GB:X03142; NID:g64575; PIDN:CAA26914.1; PID:g64576  
A>Note: initiator Met not shown  
C:Genetics:  
C:Introns: 31/2; 105/3  
C:Superfamily: Globin; globin homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxygen carrier  
F:4-147/Domain: globin homology <GLB>  
F:64/Binding site: oxygen (His) (proximal axial ligand) #status predicted  
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 66.7%; Score 26; DB 2; Length 147;  
Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
:||||  
21 VEHDG 25

Db 21 VEHDG 25

RESULT 39  
A69137  
hypothetical protein MTH291 - Methanobacterium thermoautotrophicum (strain Delta H)  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
A:Accession: A69137  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; F  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi  
A:Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: A69137  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-148 <MTH>  
A:Cross-references: GB:AE000814; GB:AE000666; NID:g2621334; PIDN:AA84797.1; PID:g262134  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH291  
C:Superfamily: hypothetical protein MTH291

Query Match 66.7%; Score 26; DB 1; Length 148;  
Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy      2 EHDGI 6
Db      71 DHDGI 75

RESULT 40
C70763
hypothetical protein Rvi558 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: C70763
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70763
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-148 <COL>
A:Cross-references: GB:274020; GB:AL123456; NID:G3261584; PIDN:CAA98333.1; PID:G1403484
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rvi558

Query Match      66.7%; Score 26; DB 2; Length 148;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LEHDG 5
Db      59 VEHDG 63

RESULT 41
A84358
hypothetical protein Vng2081h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84358
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil, Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: A84358
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-61 <STO>
A:Cross-references: GB:AE004437; NID:G10581498; PIDN:AKG20229.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG2081H

Query Match      64.1%; Score 25; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 EHDG 5
Db      33 EHDG 36

RESULT 42
H86661
hypothetical protein ycjF [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86661

R:Polotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp.
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86661
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-66 <STO>
A:Cross-references: GB:AE005176; PID:G12723159; PIDN:AAK04394.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ycjF

Query Match      64.1%; Score 25; DB 2; Length 66;
Best Local Similarity 86.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 EHDGIN 7
Db      17 EFDGVN 22

RESULT 43
T17015
metallothionein-like protein AMT2 - apple tree
C:Species: Malus domestica (apple tree)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17015
R:Reid, S.J.; Ross, G.S.
Physiol. Plantarum 100, 183-189, 1997
A>Title: Up-regulation of two cDNA clones encoding metallothionein-like proteins in app
A:Reference number: Z18652
A:Accession: T17015
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-66 <REL>
A:Cross-references: EMBL:U61974; NID:G1655852; PID:G1655853
C:Genetics:
A:Gene: AMT2
C:Keywords: metal binding

Query Match      64.1%; Score 25; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 EHDG 5
Db      46 EHDG 49

RESULT 44
G82027
hypothetical protein NMA0321 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: G82027
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jageis, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: G82027
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-68 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:G7378778; PIDN:CAB83626.1; PID:G737907
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0321

Query Match      64.1%; Score 25; DB 2; Length 68;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 EHDGI 6  
Db 21 KHDGI 25

RESULT 45

B90762  
hypothetical protein ECs1066 [imported] - Escherichia coli (strain O157:H7, substrain R1  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: B90762  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: B90762  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-72 <RAY>  
A;Cross-references: GB:BA000007; PIDN:BA34489.1; PID:gl3360526; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain R1MD 050952  
C;Genetics:  
A;Gene: ECs1066

Query Match 64.1%; Score 25; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred.No. 2.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDG 5  
Db 65 EHDG 68

Search completed: June 15, 2004, 12:25:54  
Job time : 22 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2004, 12:22:52 ; Search time 11 seconds  
(without alignments)  
33.136 Million cell updates/sec

Title: US-09-394-019c-212

Perfect score: 39

Sequence: 1 LEHDGIN 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 30433

Minimum DB seq length: 0

Maximum DB seq length: 150

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	79.5	61	1 XYLU_ECOLI	P05056 escherichia
2	29	74.4	98	1 CH10_BHRSE	O32605 ehrlichia s
3	29	74.4	130	1 DSRE_CHRYI	O87896 chromatium
4	28	71.8	128	1 E314_ADE02	P03249 human adeno
5	28	71.8	148	1 Y105_FOWPV	Q915B5 fowlpox vir
6	27	69.2	140	1 U420_HSVNG	Q05103 marek's dis
7	26	66.7	58	1 R21B_ANASP	Q8YKX5 arabidopsis
8	26	66.7	112	1 R130_ARATH	O915a3 arabidopsis
9	26	66.7	146	1 HB22_YENLA	P02133 xenopus lae
10	26	66.7	148	1 YF58_MYCTU	Q10772 mycobacteri
11	25	64.1	65	1 MT3_MUSAC	Q40256 musa acumin
12	25	64.1	66	1 MT3_MALDO	O24059 malus domes
13	25	64.1	66	1 SCXL_ANDAU	P80950 androctonus
14	25	64.1	80	1 IFIC_PSINU	Q8WHY8 psilotom nu
15	25	64.1	91	1 XIP_HUMAN	O43504 homo sapien
16	25	64.1	122	1 SUPK_ECOLI	P77667 escherichia
17	25	64.1	127	1 CMGF_BACSU	P25958 bacillus su
18	25	64.1	128	1 GC5H_NEIMA	Q91VPI neisseria m
19	25	64.1	138	1 YL75_PYRAE	Q8ZVQ1 pyrobaculum
20	25	64.1	140	1 GVPO_HALME	Q02240 halobacteri
21	25	64.1	142	1 NDK_ECOLI	Q8FF53 escherichia
22	25	64.1	142	1 NDK_ECOLI	P24233 escherichia
23	25	64.1	142	1 NDK_SALTY	Q8XFN4 salmoneilla
24	25	64.1	144	1 Y991_METUA	Q58398 methanococc
25	25	64.1	146	1 KLA2_ECOLI	P52603 escherichia
26	25	64.1	146	1 PPDD_ECOLI	P36647 escherichia
27	24	61.5	60	1 YH05_VACCV	P17356 vaccinia vi
28	24	61.5	63	1 YDIE_ECOLI	P40721 escherichia
29	24	61.5	69	1 Y082_BACAN	Q95MV3 bacillus an
30	24	61.5	73	1 LGEN_SOVIN	P59761 bos taurus
31	24	61.5	74	1 Z123_HUMAN	P35273 homo sapien
32	24	61.5	86	1 RNPH_PSEST	P57111 pseudomonas
33	24	61.5	95	1 RL25_HAEIN	P54281 haemophilus

34 24 61.5 95 1 RS6\_CLOAB Q97CX2 clostridium  
35 24 61.5 97 1 CMGG\_BACLI Q8VQ71 bacillus li  
36 24 61.5 99 1 PYY\_DICLA Q9PT99 dicentrarch  
37 24 61.5 106 1 FER2\_CAUCR P37098 caulobacter  
38 24 61.5 116 1 GCAD\_BACCL P42617 bacillus ca  
39 24 61.5 119 1 FOLB\_STRCO Q9X810 streptomyce  
40 24 61.5 120 1 PHNA\_HAEIN P44479 haemophilus  
41 24 61.5 128 1 YD2P\_BACSU P66573 bacillus su  
42 24 61.5 132 1 MRXF\_KLEPN P21650 klebsiella  
43 24 61.5 132 1 RS12\_BRYGR O59936 erysiphe gr  
44 24 61.5 133 1 SPAK\_SHIFL P35530 shigella fl  
45 24 61.5 133 1 SPAK\_SHISO Q55296 sh-gellia so

ALIGNMENTS

RESULT 1  
XYLU\_ECOLI STANDARD; PRT; 61 AA.  
ID XYLU\_ECOLI AC P05056;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Very hypothetical xylu protein.  
GN XYLU.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE=87016348; PubMed=3532033;  
RA Kurose N., Watanabe K., Kimura A.;  
RT "Nucleotide sequence of the gene responsible for D-xylose uptake in  
RT Escherichia coli.";  
RL Nucleic Acids Res. 14:7115-7123 (1986).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN=K12 / MGI655;  
RX MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
RA Oshima T., Oyama S., Saico N., Sampei G., Satoh Y., Sivasundaram S.,  
RA Tagami H., Takanashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features.";  
RNA Res. 4:91-113 (1997).  
CC -!- CAUTION: This sequence, according to the EcoSeq database (K. Rudd)  
CC is probably not a real protein; therefore this entry will probably  
CC be deleted in future releases.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; X04387; CAA27976.1; -;  
DR EMBL; D90866; BAA16222.1; -;  
DR F01; A24334; UFECC;  
KW Hypothetical protein.  
SQ SEQUENCE 61 AA; 6305 MW; 398B704F1B7CA3CA CRC64;

Query Match 79.5%; Score 31; DB 1; Length 61;  
Best Local Similarity 71.4%; Pred. No. 5.2;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHGGIN 7  
Db 36 LOHGGIS 42

RESULT 2  
ID CH10\_EHRS SEQUENCE FROM N.A. PRT; 98 AA.  
AC 032605; STANDARD; PRT; 98 AA.  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 10 kDa chaperonin (Protein Cpn10) (GroES protein).  
GN GROS OR GROES.  
OS Enrichia sennetsu.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Neorickettsia.  
OX NCBI\_TaxID=951;  
RN SEQUENCE FROM N.A.  
RP STRAIN=Japan;  
RC MEDLINE=97359489; PubMed=9215585;  
RA Zhang Y., Ohashi N., Lee E.H., Tamura A., Rikhiya Y.;  
RT "Enrichia sennetsu groE operon and antigenic properties of the GroEL homolog.";  
RL FEMS Immunol. Med. Microbiol. 18:39-46(1997).  
CC -!- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses the ATPase activity of the latter.  
CC -!- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to the groES chaperonin family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; U88092; AAB64089.1; -;  
HSSP; P05380; 1AON.  
DR HAMAP; MF\_00580; -; 1.  
DR InterPro; IPR001476; Chaprinin\_Cpn10.  
DR Pfam; PF00166; cpn10; 1.  
DR PRINTS; PR00297; CHAPERONIN10.  
DR PRODom; PD000566; Chaprinin\_Cpn10; 1.  
DR PROSITE; PS00681; CHAPERONINS\_CPN10; FALSE\_NEG.  
KW Chaperone.  
SQ SEQUENCE 98 AA; 10621 MW; 4E566F4E9D7A0CA8 CRC64;

Query Match 74.4%; Score 29; DB 1; Length 98;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHGGI 6  
Db 76 VEHDGV 81

RESULT 3  
ID DSRE\_CHRVI SEQUENCE FROM N.A. PRT; 130 AA.  
AC 087836; STANDARD; PRT; 130 AA.  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Intracellular sulfur oxidation protein dsrE.  
GN DSRE.  
OS Chromatium vinosum.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;  
OC Chromatiaceae; Allochromatium.

OX NCBI\_TaxID=1049;  
RN SEQUENCE FROM N.A.  
RP STRAIN=D / ATCC 17899 / DSM 180;  
RC MEDLINE=98361034; PubMed=9695921;  
RA Pott A.S., Dani C.;  
RT "Sirohaem sulfite reductase and other proteins encoded by genes at the dsr locus of Chromatium vinosum are involved in the oxidation of intracellular sulfur.";  
RL Microbiology 144:1881-1894(1998).  
CC -!- FUNCTION: NOT KNOWN. COULD BE INVOLVED IN THE OXIDATION OF INTRACELLULAR SULFUR.  
CC -!- SIMILARITY: Belongs to the UPF0163 (dsrE) family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; U84760; AAC35396.1; -;  
DR HAMAP; MF\_00390; -; 1.  
DR InterPro; IPR003787; DsrE.  
DR Pfam; PF02635; DsrE; 1.  
SQ SEQUENCE 130 AA; 14588 MW; 1429B54A99211B94 CRC64;

Query Match 74.4%; Score 29; DB 1; Length 130;  
Best Local Similarity 80.0%; Pred. No. 33;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
Db 41 HDGVN 45

RESULT 4  
E314\_ADE02  
ID E314\_ADE02 STANDARD; PRT; 128 AA.  
AC F03249;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Early E3B 14 kDa protein.  
OS Human adenovirus type 2, and  
OS Human adenovirus type 6.  
OS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=10515, 10534;  
RN SEQUENCE FROM N.A.  
RP SPECIES=Human adenovirus type 2;  
RC MEDLINE=81198965; PubMed=6262722;  
RA Herisse J., Galibert F.;  
RT "Nucleotide sequence of the EcoRI E fragment of adenovirus 2 genome.";  
RL Nucleic Acids Res. 9:1229-1240(1981).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human adenovirus type 6;  
RA Reichmann H., Schaarschmidt E., Gaisler B., Hausmann J., Ortman D., Bauer U., Flunker G., Seidel W.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP IDENTIFICATION OF PROTEIN.  
RX MEDLINE=90156523; PubMed=2304142;  
RA Horton T.H., Tolleson A.E., Wold W.S.M., Gooding L.R.;  
RT "A protein serologically and functionally related to the group C E3 14,700-kilodalton protein is found in multiple adenovirus serotypes.";  
RL J. Virol. 64:1250-1255(1990).  
CC -!- FUNCTION: Protects virus-infected cells from TNF-induced cytotoxicity.  
CC -!- SIMILARITY: TO EQUIVALENT PROTEIN IN OTHER ADENOVIRUSES.

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01917; ; NOT ANNOTATED_CDS.
DR EMBL; V16037; CA875954.1; -.
DR PIR; A03818; ERA042.
DR InterPro; IPR004985; Adeno_E3_153.
DR Pfam; PF03307; Adeno_E3_15_3; 1.
KW Early protein.
SQ SEQUENCE 128 AA; 14738 MW; 09882D90B253F146 CRC64;
Query Match 71.8%; Score 28; DB 1; Length 128;
Best Local Similarity 85.7%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LEHDGIN 7
DB 7 LELDGIN 13
RESULT 5
V105_FOWPV STANDARD; PRT; 148 AA.
AC Q9J5B5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein FFW105.
GN FFW105.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus."
RL J. Virol. 74:3815-3831(2000).
CC -!- SIMILARITY: BELONGS TO THE POXVIRUSES F15 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF198100; AAF44449.1; -.
DR InterPro; IPR007675; Pox_F15.
DR Pfam; PF04596; Pox_F15; 1.
DR PRSF; PRSF015694; VAC_F15L; 1.
SQ SEQUENCE 148 AA; 17723 MW; 492B0CB93D6B02A CRC64;
Query Match 71.8%; Score 28; DB 1; Length 148;
Best Local Similarity 66.7%; Pred. No. 63;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 EHDGIN 7
DB 111 EHNLGN 116
RESULT 6
U420_HSVMG STANDARD; PRT; 140 AA.
ID U420_HSVMG
```

```
AC Q05103;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 15.5 kDa protein.
GN U5420.
OS Marek's disease herpesvirus (strain GA) (MDHV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=10388;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=931118245; PubMed=1282282;
RA Sakaguchi M., Urakawa T., Hirayama Y., Miki N., Yamamoto M.,
RA Hirai K.;
RT "Sequence determination and genetic content of an 8.9-kb restriction
RT fragment in the short unique region and the internal inverted repeat
RT of Marek's disease virus type 1 DNA."
RL Virus Genes 6:365-378(1992).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80595; AAB59889.1; -.
KW Hypothetical protein.
SQ SEQUENCE 140 AA; 15496 MW; C3A73F42C3CA6545 CRC64;
Query Match 69.2%; Score 27; DB 1; Length 140;
Best Local Similarity 71.4%; Pred. No. 98;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LEHDGIN 7
DB 35 LLEHGIN 41
RESULT 7
R21B_ANASP STANDARD; PRT; 58 AA.
ID R21B_ANASP
AC Q8YRK5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 30S ribosomal protein S21.B.
GN RPS22 OR RPS21 OR ASR0843.
OS Anabaena sp. (Strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Belongs to the S21P family of ribosomal proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC EMBL; AP003593; BAB72800.1; -  
 DR PIR; A11911; A11911.  
 DR HAMAP; MF 00358; -; 1.  
 DR InterPro; IPR001911; Ribosomal\_S21.  
 DR Pfam; PF01165; Ribosomal\_S21.  
 DR PRINTS; PR00976; Ribosomal\_S21.  
 DR ProDom; PD005521; Ribosomal\_S21; 1.  
 DR TIGRfam; TIGR00030; S21p1.  
 DR PROSITE; PS01181; Ribosomal\_S21; 1.  
 DR Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 58 AA; 6901 MW; 5BD5B2B6A93A9B4D CRC64;  
 Query Match 66.7%; Score 26; DB 1; Length 58;  
 Best Local Similarity 66.7%; Pred. No. 60;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGIN 7  
 DB 8 EHEGID 13

RESULT 8  
 RL30 ARATH STANDARD; PRT; 112 AA.  
 ID RL30 ARATH  
 AC Q9LSA3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DR InterPro; IPR000231; Ribosomal\_L30e.  
 DR Pfam; PF01248; Ribosomal\_L30e; 1.  
 DR ProDom; PD004495; Ribosomal\_L30e; 1.  
 DR PRINTS; PR00709; Ribosomal\_L30e; 1.  
 DR PROSITE; PS00993; Ribosomal\_L30e; 1.  
 DR Ribosomal protein.  
 KW Ribosomal protein.  
 SQ SEQUENCE 112 AA; 12279 MW; BLB2286ED2CC0549 CRC64;  
 Query Match 66.7%; Score 26; DB 1; Length 112;  
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HDGIN 7  
 DB 11 HEGIN 15

CC EMBL; X03142; CAA26914.1; -  
 DR EMBL; M21411; AAA49653.1; -  
 DR PIR; A54492; A54492.  
 DR HSP; P02100; 1A9W.  
 DR InterPro; IPR002337; Beta haem.  
 DR InterPro; IPR000971; Globin.  
 DR Pfam; PF00042; Globin; 1.  
 DR PRINTS; PR00814; BETAHEM.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 DR Heme; Oxygen transport; Transport; Erythrocyte.  
 KW INIT MET 0  
 FT METAL 63 63 IRON (HEME DISTAL LIGAND).  
 FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).  
 SQ SEQUENCE 146 AA; 15841 MW; BB53347CA9BFFAE CRC64;  
 Query Match 66.7%; Score 26; DB 1; Length 146;  
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LEHDG 5  
 DB 20 VEHDG 24

RESULT 10  
 YF58 MYCTU STANDARD; PRT; 148 AA.  
 ID YF58 MYCTU  
 AC Q10772;  
 DT 01-OCT-1996 (Rel. 34, Created)

RESULT 9  
 HBB2 XENLA STANDARD; PRT; 146 AA.  
 ID HBB2 XENLA  
 AC P02133; P06644;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DR InterPro; IPR001911; Ribosomal\_S21.  
 DR Pfam; PF01165; Ribosomal\_S21.  
 DR PRINTS; PR00976; Ribosomal\_S21.  
 DR ProDom; PD005521; Ribosomal\_S21; 1.  
 DR TIGRfam; TIGR00030; S21p1.  
 DR PROSITE; PS01181; Ribosomal\_S21; 1.  
 DR Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 58 AA; 6901 MW; 5BD5B2B6A93A9B4D CRC64;  
 Query Match 66.7%; Score 26; DB 1; Length 58;  
 Best Local Similarity 66.7%; Pred. No. 60;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGIN 7  
 DB 8 EHEGID 13

RESULT 10  
 YF58 MYCTU STANDARD; PRT; 148 AA.  
 ID YF58 MYCTU  
 AC Q10772;  
 DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DT Hypothetical protein RV1558/MT1609/MD1584.  
GN RV1558 OR MT1609 OR MTCV48.07C OR MB1584.  
OS Mycobacterium tuberculosis, and  
OC Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773, 1765;  
RN [1]  
RN  
SEQUENCE FROM N.A.  
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544 (1998).  
RN [2]  
RN  
SEQUENCE FROM N.A.  
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL J. Bacteriol. 184:5479-5490 (2002).  
RN [3]  
RN  
SEQUENCE FROM N.A.  
RP SPECIES=M.bovis; STRAIN=AF2122/97;  
RX MEDLINE=22709107; PubMed=12788972;  
RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,  
RA Pryor M., Duttoy S., Gordin S., Lacroix C., Monsenpe C., Simon S.,  
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
RA "The complete genome sequence of Mycobacterium bovis.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
RL [4]  
RN  
-1- SIMILARITY: TO M.TUBERCULOSIS RV1261C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z74020; CAA98333.1; -;  
DR EMBL; AE007027; AAK45876.1; -;  
DR EMBL; BX248339; CAD96251.1; -;  
DR PIR; C70763; C70763.  
DR TIGR; MT1609; -;  
DR TubercuList; RV1558; -;  
DR InterPro; IPR04378; Mtu\_fam\_11.  
DR Pfam; PF04075; DUF385; 1.  
DR TIGRFAMs; TIGR00026; Mtu\_fam\_11; 1.  
DR Hypothetical protein; Complete proteome.  
KW SEQUENCE 148 AA; 16347 MW; A9F98E9039871938 CRC64;  
SQ  
Query Match 66.7%; Score 26; DB 1; Length 148;  
Best Local Similarity 80.0%; Pred No. 1.7e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LEHDG 5

Db 59 VEHDG 63  
|||||  
ID MT3\_MUSAC STANDARD; PRT; 65 AA.  
AC Q40256;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Metallothionein-like protein type 3 (MT-3).  
OS Musa acuminata (Banana).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;  
OC Musa.  
OX NCBI\_TaxID=4641;  
RN [1]  
RN  
SEQUENCE FROM N.A.  
RP STRAIN=cv. Grand nain;  
RX MEDLINE=98002325; PubMed=9342866;  
RA Clendennen S.K., May G.D.;  
RA "Differential gene expression in ripening banana fruit.";  
RT Plant Physiol. 115:463-469 (1997).  
RL  
CC -1- FUNCTION: Metallothioneins have a high content of cysteine  
CC residues that bind various heavy metals.  
CC  
CC -1- SIMILARITY: Belongs to the metallothionein superfamily; family 15.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U49044; BAB82615.1; -;  
DR InterPro; IPR00347; Metallothion\_15.  
DR Pfam; PF01439; Metallothion\_2; 1.  
DR ProDom; PD001611; Metallothion\_15; 1.  
KW Metal-binding; Metal-thiolate Cluster.  
SQ SEQUENCE 65 AA; 6774 MW; F456C0668AB4B577 CRC64;  
Query Match 64.1%; Score 25; DB 1; Length 65;  
Best Local Similarity 100.0%; Pred No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 EHDG 5  
|||||  
Db 45 EHDG 48  
ID MT3\_MALDO STANDARD; PRT; 66 AA.  
AC O24059;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Metallothionein-like protein type 3 (MT-3).  
OS Malus domestica (Apple) (Malus sylvestris).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
OX NCBI\_TaxID=3750;  
RN [1]  
RN  
SEQUENCE FROM N.A.  
RA Reid S.J., Ross G.S.;  
RT "Up-regulation of two cDNA clones encoding metallothionein-like  
RT proteins in apple fruit during cool storage.";  
RL Physiol. Plantarum 100:183-189 (1997).  
RL  
CC -1- FUNCTION: Metallothioneins have a high content of cysteine

```
CC residues that bind various heavy metals.
CC -!- SIMILARITY: Belongs to the metallothionein superfamily; family 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U61974; AAC23698.1; -.
CC PIR; T17015; T17015.
CC InterPro; IPR000347; Metallothion_15.
CC ProDom; PD001611; Metallothion_15; 1.
CC Metal-binding; Metal-thiolate Cluster.
CC KW SEQUENCE 66 AA; 6930 MW; BFCF793C36FC036E CRC64;
CC
CC Query Match 64.1%; Score 25; DB 1; Length 66;
CC Best Local Similarity 100.0%; Pred. No. 1.1e+02;
CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 2 EHDG 5
CC Db 46 EHDG 49
CC
CC RESULT 13
CC SCXL_ANDAU STANDARD; PRT; 66 AA.
CC ID SCXL_ANDAU
CC AC P80950;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Neurotoxin-like protein STR1 (Anatoxin AaH STR1) (AaHSTR1).
CC OS Androctonus australis (Sahara scorpion).
CC CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
CC CC Eutherozoa; Eutherozoa; Eutherozoa; Eutherozoa; Eutherozoa;
CC CC NCBI_TaxID=6858;
CC RN [1]
CC RP SEQUENCE.
CC RC STRAIN=Hector; TISSUE=Venom;
CC RA Mansuete P., Hassani O., Cestele S., Loret E.P., van Dorselaer A.,
CC RA Rochat H., Sampieri F.;
CC RL Submitted (AUG-1997) to Swiss-Prot.
CC RN [2]
CC RP STRUCTURE BY NMR.
CC RC STRAIN=Hector;
CC RA Blanc E., Hassani O., Meunier S., Mansuelle P., Sampieri P.,
CC RA MEDLINE=97433298; PubMed=9288938;
CC RT "1H-NMR-derived secondary structure and overall fold of a natural
CC RT anatoxin from the scorpion Androctonus australis hector.";
CC RL Eur. J. Biochem. 247:1118-1126(1997).
CC CC -!- FUNCTION: This protein is not toxic.
CC CC -!- SUBCELLULAR LOCATION: Secreted.
CC CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC CC -!- SIMILARITY: Belongs to the alpha/beta-scorpion toxin family. Beta-
CC toxin subfamily.
CC HSSP; P01484; 1AHO.
CC DR InterPro; IPR003614; Knot1.
CC DR InterPro; IPR002061; Scorpion_toxinL.
CC DR Pfam; PF00537; toxin_3; 1.
CC DR ProDom; PD000908; Scorpion_toxinL; 1.
CC DR SMART; SM00505; Knot1; 1.
CC FT DISULFID 13 64
CC FT DISULFID 17 40
CC FT DISULFID 26 45
CC FT DISULFID 30 47
CC SEQUENCE 66 AA; 7641 MW; D1B64B7C4E8467E3 CRC64;
CC
CC Query Match 64.1%; Score 25; DB 1; Length 66;
CC Best Local Similarity 80.0%; Pred. No. 1.1e+02;
CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 HDGIN 7
Db 8 HDGTN 12
RESULT 14
IFIC_PSINU STANDARD; PRT; 80 AA.
ID IFIC_PSINU
AC Q8WHY8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Translation initiation factor IF-1, chloroplast.
GN INFA.
OS Psilotum nudum (Whisk fern).
CC Chloroplast.
CC CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC CC Moniliformopses; Psilotophyta; Psilotales; Psilotaceae; Psilotum.
CC CC NCBI_TaxID=3240;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Kinyoku;
CC RA Wakasugi T., Nishikawa A., Yamada K., Sugiyama M.;
CC RT "Complete nucleotide sequence of the chloroplast genome from a fern,
CC RT Psilotum nudum.";
CC RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC CC -!- FUNCTION: No specific function has so far been attributed to this
CC CC initiation factor; however, it seems to stimulate more or less all
CC CC the activities of the other two initiation factors, IF-2 and IF-3.
CC CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC CC -!- SIMILARITY: Belongs to the IF-1 family.
CC CC -!- SIMILARITY: Contains 1 S1-like domain.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004638; BAB84251.1; -.
CC DR HAMAP; MF_00075; 1.
CC DR InterPro; IPR008994; Nucleic_acid_OB.
CC DR InterPro; IPR003029; S1.
CC DR InterPro; IPR006196; S1_IP1.
CC DR InterPro; IPR004368; TIF_IP1.
CC DR Pfam; PF00575; S1; 1.
CC DR SMART; SM00316; S1; 1.
CC DR TIGRFAMs; TIGR00008; infa; 1.
CC DR PROSITE; PS00832; S1_IP1_TYPE; 1.
CC KW Initiation factor; Protein biosynthesis; Chloroplast.
CC FT DOMAIN 1 72
CC FT S1-LIKE.
CC SQ SEQUENCE 80 AA; 9299 MW; CB632F174EACDB7D CRC64;
CC
CC Query Match 64.1%; Score 25; DB 1; Length 80;
CC Best Local Similarity 83.3%; Pred. No. 1.4e+02;
CC Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 2 EHDGIN 7
CC Db 3 EHDLIN 8
CC
CC RESULT 15
CC XIP_HUMAN STANDARD; PRT; 91 AA.
CC ID XIP_HUMAN
CC AC Q43504;
CC DT 15-JUL-1999 (Rel. 38, Created)
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

DE Hepatitis B virus X interacting protein (HBX-interacting protein) (HBV)  
DE X interacting protein).  
CN HBXP OR XIP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC TISSUE=Liver.  
RX MEDLINE=98139062; PubMed=9499022;  
RA Melegari M., Scaglioni P.P., Wands J.R.;  
RT "Cloning and characterization of a novel hepatitis B virus X binding  
RT protein that inhibits viral replication.";  
RL J. Virol. 72:1737-1743(1998).  
CC -|- FUNCTION: Binds to hepatitis B virus (HBV) replication and  
CC down-regulating hepatitis B virus (HBV) replication.  
CC -|- TISSUE SPECIFICITY: Highly expressed in skeletal and cardiac  
CC muscle, followed by pancreas, kidney, liver, brain, placenta and  
CC lung.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AF029890; AAC52032.1; -;  
CC Genew; HGNC:17955; HBXP.  
CC GO; GO:0003800; F:antiviral response protein activity; TAS.  
CC GO; GO:0007125; P:invasive growth; TAS.  
CC GO; GO:0008166; P:viral replication; TAS.  
CC MUTAGEN 12 12 T->A: NO CHANGE.  
CC MUTAGEN 36 36 (TRUNCATED FORM OF HBX).  
CC  
CC SEQUENCE 91 AA; 9614 MW; 01D9E762ABC63990 CRC64;  
CC  
CC Query Match 64.1%; Score 25; DB 1; Length 91;  
CC Best Local Similarity 80.0%; Pred. NO. 1.6e+02;  
CC Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 2 EHDGI 6  
CC Db 78 KHDGI 82  
CC  
CC  
CC RESULT 16  
CC SUFA\_ECOLI  
CC ID SUFA\_ECOLI STANDARD; PRT; 122 AA.  
CC AC P77667;  
CC DT 01-NOV-1997 (Rel. 35, Created)  
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
CC DE SuFa protein.  
CC GN SUFA OR B1684.  
CC OS Escherichia coli.  
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
CC OC Enterobacteriaceae; Escherichia.  
CC OX NCBI\_TaxID=562;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=K12 / MG1655;  
CC RX MEDLINE=97426617; PubMed=9278503;  
CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
CC RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
CC RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
CC RA Mau B., Shao Y.;  
CC RT "The complete genome sequence of Escherichia coli K-12.";  
CC Science 277:1453-1474(1997).  
CC RL [2]  
CC RN SEQUENCE FROM N.A.  
CC RP

RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sampaio G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [3]  
RN GENE NAME.  
RP STRAIN=K12 / MG1655;  
RX MEDLINE=99255563; PubMed=10322040;  
RA Patzer S.I., Hantke K.;  
RT "Sufs is a Nifs-like protein, and SufD is necessary for stability of  
RT the 2Pe-2S Phuf protein in Escherichia coli.";  
RL J. Bacteriol. 181:3307-3309(1999).  
CC -|- SIMILARITY: BELONGS TO THE HESB/YADR/YFHF FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AB000263; AAC74754.1; -;  
CC EMBL; D90811; BAA15453.1; -;  
CC EMBL; D90812; BAA15459.1; -;  
CC PIR; D64926; D64926.  
CC Ecogen; EGI1378; sufa.  
CC InterPro; IPR000361; HESB\_yadr\_yfhf.  
CC Pfam; PF01521; HESB-like; 1.  
CC ProDom; PD002183; HESB\_yadr\_yfhf; 1.  
CC TIGRFAMs; TIGR00049; TIGR00049; 1.  
CC PROSITE; PS01152; HESB; 1.  
CC Complete proteome.  
CC SEQUENCE 122 AA; 13300 MW; D26917F42B4B740B CRC64;  
CC  
CC Query Match 64.1%; Score 25; DB 1; Length 122;  
CC Best Local Similarity 100.0%; Pred. NO. 2.3e+02;  
CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 2 EHDG 5  
CC Db 71 EHDG 74  
CC  
CC  
CC RESULT 17  
CC CMGF\_BACSU  
CC ID CMGF\_BACSU STANDARD; PRT; 127 AA.  
CC AC P25958; O32021.  
CC DT 01-MAY-1992 (Rel. 22, Created)  
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)  
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
CC DE ComG operon protein 6.  
CC GN COMGF OR COMG6 OR BSU24680.  
CC OS Bacillus subtilis.  
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
CC OX NCBI\_TaxID=1423;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=9008773; PubMed=2507524;  
CC RA Albano M., Breitling R., Dubnau D.A.;  
CC RT "Nucleotide sequence and genetic organization of the Bacillus  
CC subtilis comG operon.";  
CC RL J. Bacteriol. 171:5386-5404(1989).  
CC RN [2]  
CC RN SEQUENCE FROM N.A.  
CC RP STRAIN=168;  
CC RC

RX MEDLINE=9804033; PubMed=9384377;  
 RA Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriello R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.M., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dutertre A., Ehrlich S.D., Emmerson P.T.,  
 RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
 RA Gim S.Y., Glaser P., Goffeau A., Gollightly B.J., Grandi G.,  
 RA Guiseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karanata D., Kasartha Y., Koningsstein G., Krogh S., Kumano M.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prescena E., Fujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin M., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzmeger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RA "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.";  
 RT Nature 390:249-256 (1997).  
 RN [3]  
 RN FUNCTION.  
 RP MEDLINE=9803053; PubMed=9422590;  
 RX Chung Y.S., Dubnau D.A.;  
 RA "All seven comG open reading frames are required for DNA binding  
 RT during transformation of competent Bacillus subtilis.";  
 RT J. Bacteriol. 180:41-45 (1998).  
 RN [4]  
 RN SUBCELLULAR LOCATION.  
 RP MEDLINE=98389321; PubMed=9723928;  
 RX Chung Y.S., Braid F., Dubnau D.A.;  
 RA "Cell surface localization and processing of the ComG proteins,  
 RT required for DNA binding during transformation of Bacillus subtilis.";  
 RT Mol. Microbiol. 29:905-913 (1998).  
 CC -!- FUNCTION: Required for transformation and DNA binding.  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M29691; AAA83372.1; ALT INIT.  
 CC EMBL; Z99116; CAB14399.1; --  
 CC PIR; G30338; G30338.  
 CC PIR; G69603; G69603.  
 CC Subtilisin; BGI0488; comGF.  
 KW Competence; Membrane; Complete proteome.  
 SQ SEQUENCE 127 AA; 14281 MW; 6598AA28873AD84F CRC64;  
 Query Match 64.1%; Score 25; DB 1; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDG 5  
 DB 20 EHDG 23

RESULT 18  
 GCSH NEIMA  
 ID GCSH NEIMA STANDARD; PRT; 128 AA.  
 AC Q9JVE1.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glycine cleavage system H protein.  
 DE GCVH OR NNA0759.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=2022556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Deakin K., Feltwell T., Hamlin N., Holroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RA "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491.";  
 RT Nature 404:502-506 (2000).  
 CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of  
 CC glycine. The H protein shuttles the methylamine group of glycine  
 CC from the P protein to the T protein (By similarity).  
 CC -!- COPACITOR: The H chain contains a covalently-bound lipoyl cofactor  
 CC (By similarity).  
 CC -!- SUBUNIT: The glycine cleavage system is composed of four proteins:  
 CC P, T, L and H (By similarity).  
 CC -!- SIMILARITY: Belongs to the gcvH family.  
 CC -!- SIMILARITY: Contains 1 lipoyl-binding domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AL162754; CAB804042.1; --  
 CC PIR; H81919; H81919.  
 CC HSP; P16048; 1HTP.  
 CC HAWAP; MF\_00272; --; 1.  
 CC InterPro; IPR002930; GCV\_H.  
 CC InterPro; IPR003016; Lipoyl\_Bs.  
 CC Pfam; PF01597; GCV\_H; 1.  
 CC TIGRPFAMs; TIGR00527; gcvH; 1.  
 CC PROSITE; PS00189; LIPOYL; 1.  
 KW Lipoyl; Complete proteome.  
 FT BINDING 66 66 LIPOYL (BY SIMILARITY).  
 SQ SEQUENCE 128 AA; 13640 MW; 065F4ED818BDC74 CRC64;  
 Query Match 64.1%; Score 25; DB 1; Length 128;  
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LEHDGI 6  
 DB 20 LEEDGI 25  
 RESULT 19  
 YL75 PYRAE  
 ID YL75 PYRAE STANDARD; PRT; 138 AA.  
 AC Q8ZVQ3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)



DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein PF02175  
 GN PAE2175  
 OS Pyrobaculum aerophilum  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 CC -!- SIMILARITY: Belongs to the UPF0217 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB009861; AAL64003.1; -  
 DR HAMAP; MF 00587; atypical; 1.  
 DR InterPro: IPR007158; DUF358.  
 DR Pfam; PF04013; DUF358; 1.  
 KW Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 138 AA; 15818 MW; DB86C30BA221619 CRC64;  
 Query Match 64.1%; Score 25; DB 1; Length 138;  
 Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGIN 7  
 DB :||||:  
 95 DHDGLS 100  
 RESULT 20  
 GVPO HALME STANDARD; PRT; 140 AA.  
 ID GVPO HALME STANDARD; PRT; 140 AA.  
 AC Q02240;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE Gvpo protein.  
 GN GVPO.  
 OS Halobacterium mediterranei (Haloflex mediterranei).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Haloflex.  
 OX NCBI\_TaxID=2252;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1411;  
 RX MEDLINE=93021102; PubMed=1404376;  
 RA Englert C., Krueger K., Offner S., Pfeifer F.;  
 RT "Three different but related gene clusters encoding gas vesicles in  
 RT halophilic archaea";  
 RL J. Mol. Biol. 227:586-592(1992).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X64701; CAA45942.1; -  
 DR  
 PIR; S28113; S28113.  
 DR InterPro: IPR008634; Gvpo.  
 DR Pfam; PF05800; Gvpo; 1.  
 KW Gas vesicle.  
 SQ SEQUENCE 140 AA; 15306 MW; B97A2A43E3B56BF4 CRC64;  
 Query Match 64.1%; Score 25; DB 1; Length 140;  
 Best Local Similarity 80.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 HDGIN 7  
 DB :||||:  
 11 HDGID 15  
 RESULT 21  
 NDK ECOL6 STANDARD; PRT; 142 AA.  
 ID NDK ECOL6 STANDARD; PRT; 142 AA.  
 AC Q8FF53;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)  
 DE (Nucleoside-2-P kinase).  
 GN NDK OR C3041.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:HI / CFT073 / ATCC 700928;  
 RX MEDLINE=23388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -!- FUNCTION: Major role in the synthesis of nucleoside triphosphates  
 CC other than ATP.  
 CC -!- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +  
 CC nucleoside triphosphate.  
 CC -!- SUBUNIT: Homotrimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the NDK family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE016764; AAN81491.1; -  
 DR HAMAP; MF 00451; -; 1.  
 DR InterPro: IPR001564; NDK.  
 DR Pfam; PF00334; NDK; 1.  
 DR PRINTS; PR01243; NUOCPKINASE.  
 DR PRODOM; PD001018; NDK; 1.  
 DR SMART; SM00562; NDK; 1.  
 DR PROSITE; PS00469; NDP\_KINASES; 1.  
 KW Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.  
 FT INIT MET 0 BY SIMILARITY.  
 FT ACT\_SITE 116 116 BY SIMILARITY.  
 FT MOD\_RES 118 118 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 120 120 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 142 AA; 15305 MW; DFB5C0D7A842379 CRC64;  
 Query Match 64.1%; Score 25; DB 1; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDG 5  
|  
|  
|  
|  
52 EHDG 55

Db

RESULT 22  
NDK\_ECOLI  
ID\_NDK\_ECOLI STANDARD; PRT; 142 AA.  
AC P24233;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)  
DE (Nucleoside-2-P kinase).  
GN NDK OR B9518 OR Z3781 OR ECS3380 OR SF2564 OR S2736.  
OS Escherichia coli.  
OS Escherichia coli, O157:H7, and  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 83334, 623;  
[1]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=K12;  
RC MEDLINE=92039015; PubMed=1657712;  
RX Hama H., Almaula N., Lerner C.G., Inouye S., Inouye M.;  
RA "Nucleoside diphosphate kinase from Escherichia coli; its  
RT overproduction and sequence comparison with eukaryotic enzymes.";  
RL Gene 105:31-36(1991).  
[2]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=K12 / MG1655;  
RC MEDLINE=97428617; PubMed=9278503;  
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
[3]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=K12;  
RC MEDLINE=97349980; PubMed=9205937;  
RX Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
- K12 genome corresponding to 50.0-68.8 min on the linkage map and  
analysis of its sequence features.";  
RL DNA Res. 4:91-113(1997).  
[4]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RC MEDLINE=21074935; PubMed=11206551;  
RX Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai C., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
[5]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
RC MEDLINE=21156231; PubMed=11258796;  
RX Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
[6]  
RN SEQUENCE OF 1-45.  
RC SPECIES=E.coli;  
RX MEDLINE=92362321; PubMed=1323446;  
RA Ray N.B., Mathews C.K.;  
RT "Nucleoside diphosphokinase: a functional link between intermediary  
RT metabolism and nucleic acid synthesis.";  
RL Curr. Top. Cell. Regul. 33:343-357(1992).  
[7]  
RN CHARACTERIZATION, AND PARTIAL SEQUENCE.  
RC SPECIES=E.coli;  
RX MEDLINE=95247689; PubMed=7730286;  
RA Almaula N., Lu Q., Delgado J., Belkin S., Inouye M.;  
RT "Nucleoside diphosphate kinase from Escherichia coli.";  
RL J. Bacteriol. 177:2524-2529(1995).  
[8]  
RN SEQUENCE OF 1-11.  
RC SPECIES=E.coli; STRAIN=K12 / W3110;  
RX Frutiger S., Hughes G.J., Pasquali C., Hochstrasser D.F.;  
RN Submitted (FEB-1996) to Swiss-Prot.  
[9]  
RN SEQUENCE FROM N.A.  
RP SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RC MEDLINE=2272406; PubMed=12384590;  
RX Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
RT through comparison with genomes of Escherichia coli K12 and O157";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
[10]  
RN SEQUENCE FROM N.A.  
RP SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RC MEDLINE=22590274; PubMed=12704152;  
RX Wei J., Goldberg M.B., Burland V., Vankatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
flexneri serotype 2a strain 2457T";  
RL Infect. Immun. 71:2775-2786(2003).  
CC -1- FUNCTION: Major role in the synthesis of nucleoside triphosphates  
CC other than ATP.  
CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +  
CC nucleoside triphosphate.  
CC -1- SUBUNIT: Homotetramer.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to the NDK family.  
[11]  
RN This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
[12]  
RN EMBL; X57555; CAA40780.1; -;  
DR EMBL; AE000338; AAC75571.1; -;  
DR EMBL; D90881; BAA16405.1; -;  
DR EMBL; D90882; BAA16409.1; -;  
DR EMBL; AE005481; AAG57628.1; -;  
DR EMBL; AP002561; BAB36803.1; -;  
DR EMBL; AE015272; AAN40404.1; -;  
DR EMBL; AE016987; AAP17891.1; -;  
DR PIR; D91051; D91051.

```

DR PIR; H5895; H85895.
DR PIR; JH0495; JH0495.
DR HSP; P15266; INHK.
DR HSP; P24233; -.
DR SWISS-2DPAGE; P24233; COLI.
DR Ecogene; E010650; ndk.
DR HAMAP; MF_00451; -.
DR InterPro; IPR001564; NDK.
DR Pfam; PF00334; NDK; 1.
DR PRINTS; PRO1243; NUCDPKINASE.
DR ProDom; PD001018; NDK; 1.
DR SMART; SM00562; NDK; 1.
DR PROSITE; PS00469; NDP_KINASES; 1.
KW Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT INIT MET 0
FT ACT SITE 116 116
FT MOD_RES 118 118 PHOSPHORYLATION.
FT MOD_RES 120 120 PHOSPHORYLATION.
SQ SEQUENCE 142 AA; 15332 MW; DFB7AC107A842364 CRC64;

Query Match 64.1%; Score 25; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDG 5
DB 52 EHDG 55

RESULT 23
NDK_SALTY STANDARD; PRT; 142 AA.
AC O8XFN4,
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
DE (Nucleoside-2-P kinase).
GN NDK OR STM2526 OR STV2771 OR T0330.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;
[1]
RN SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=1677609;
RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Forwell S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
[2]
RN SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=1677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham A., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[3]
RN SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Major role in the synthesis of nucleoside triphosphates
other than ATP.
CC -!- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +
nucleoside triphosphate.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the NDK family.
-----
This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE008814; AAL21420.1; -.
EMBL; AL627275; CAD02729.1; -.
EMBL; AE016835; AAO68053.1; -.
StyGene; SG23232?; ndk.
HAMAP; MF_00451; -.
InterPro; IPR001564; NDK.
Pfam; PF00334; NDK; 1.
PRINTS; PRO1243; NUCDPKINASE.
ProDom; PD001018; NDK; 1.
SMART; SM00562; NDK; 1.
PROSITE; PS00469; NDP_KINASES; 1.
Transferase; Kinase; ATP-binding; Complete proteome.
FT INIT MET 0
FT ACT SITE 116 116 BY SIMILARITY.
SQ SEQUENCE 142 AA; 15390 MW; C27B2B10BFF7F919 CRC64;

Query Match 64.1%; Score 25; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDG 5
DB 52 EHDG 55

RESULT 24
Y991_METUA STANDARD; PRT; 144 AA.
AC Q58398;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0991.
GN MJ0991.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., FitzGerald L.M., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitch J., Tomb J.-F., Adams M.D., Reich C.I.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Overbeek R., Kirkness E.F., Weinstock N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Scott J.L., Geoghagen N.S.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

```

```
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67541; AAB98996.1; -.
DR PIR; G64423; G64423.
DR TIGR; MJ0991; -.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 17270 MW; 00F1CFCD0007A09E CRC64;
Query Match 64.1%; Score 25; DB 1; Length 144;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 LEHDGIN 7
Db 79 LDYDGD 85
RESULT 25
KLA2_ECOLI
ID KLA2_ECOLI STANDARD; PRT; 146 AA.
AC P52603; Q47333;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Antirestriction protein KicA.
GN KICA OR KILC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Larsen M.H., Figurski D.H.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
[2]
RC STRAIN=K12;
RX MEDLINE=95291464; PubMed=7773415;
RA Thomas C.W., Smith C.A., Ibbotson J.P., Johnston L., Wang N.;
RT "Evolution of the kora-oriV segment of promiscuous IncP plasmids.";
RL Microbiology 141:1201-1210(1995).
CC -!- FUNCTION: COULD BE INVOLVED IN OVERCOMING RESTRICTION BARRIERS
CC DURING ESTABLISHMENT AFTER CONJUGATIVE TRANSFER.
CC -!- SIMILARITY: TO E.COLI YFUX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U05773; AAA57448.1; -.
DR EMBL; L13392; AAA75284.1; -.
DR PIR; A55857; A55857.
DR PIR; I41323; I41323.
DR InterPro; IPR004914; Antirestrict.
DR Pfam; PF03230; Antirestrict; 1.
KW Plasmid.
FT CONFLICT 24 4 V -> VMF (IN REF. 2).
L -> P (IN REF. 2).
```

```
FT CONFLICT 33 33 L -> P (IN REF. 2).
FT CONFLICT 77 77 R -> P (IN REF. 2).
SQ SEQUENCE 146 AA; 15867 MW; 9835588E26B8BA60 CRC64;
Query Match 64.1%; Score 25; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 EHDG 5
Db 73 EHDG 76
RESULT 26
PPDD_ECOLI
ID PPDD_ECOLI STANDARD; PRT; 146 AA.
AC P36647; Q8RMZ6;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Prepilin peptidase dependent protein D precursor.
GN PPDD OR B0108.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95047556; PubMed=7959070;
RA Whitchurch C.B., Mattick J.S.;
RT "Escherichia coli contains a set of genes homologous to those
RT involved in protein secretion, DNA uptake and the assembly of type-4
RT fimbriae in other bacteria.";
RL Gene 150:9-15(1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[4]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Shatia R.S.;
RL Submitted (XXY-1993) to the EMBL/GenBank/DBJ databases.
[5]
RP IDENTIFICATION.
RA Rudd K.R.;
RL Unpublished observations (JAN-1994).
CC -!- FUNCTION: Not yet known.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U28105; AAC36923.1; -.

```

```
DR EMBL; D26562; BAB96678.1; ALT_INIT.
DR EMBL; AE000119; AAC73219.1; -.
DR EMBL; D20834; -. NOT_ANNOTATED_CDS.
DR PIR; D64733; D64733.
DR HSP; P02974; 2PIL.
DR EcoGene; EG12107; ppdD.
DR InterPro; IPR001082; Pili.
DR Pfam; PF00114; pilin; 1.
DR ProDom; PD008666; Pilin; 1.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Methylation; Complete proteome.
FT PROPEP 1 6 BY SIMILARITY.
FT CHAIN 7 146 PREPILIN PEPTIDASE DEPENDENT PROTEIN D.
FT MOD_RES 7 7 METHYLATION (BY SIMILARITY).
FT CONFLICT 1 5 MDKOR -> DQOAT (IN REF. 2).
SQ SEQUENCE 146 AA; 15622 MW; 7905D78A1E8B17A0 CRC64;

Query Match 64.1%; Score 25; DB 1; Length 146;
Best Local Similarity 57.1%; Pred. No. 2.8e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7
||| |
DB 58 LEHGGLD 64

RESULT 27
YH05_VACCV STANDARD; PRT; 60 AA.
AC YH05_VACCV
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Hypothetical 5.7 kDa HindIII-C protein.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89073756; PubMed=2849238;
RA Kotwal G.J.; Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
RL vaccinia virus terminal transposition mutant.";
RL Virology 167:524-537(1988).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; M22812; AAA69598.1; -.
DR PIR; G31829; WZVZA7.
KW Hypothetical protein; Early protein.
SQ SEQUENCE 60 AA; 5680 MW; B29BDE019384BEB CRC64;

Query Match 61.5%; Score 24; DB 1; Length 60;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HDGIN 7
||| |
DB 19 HDGSN 23

RESULT 28
YDIE_ECOLI STANDARD; PRT; 63 AA.
AC P40721;
```

```
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydIE.
GN ydIE OR B1705 OR Z2734 OR ECS2412.
OS Escherichia coli, and
OS Escherichia coli, O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91323737; PubMed=1677907;
RA Hudson G.S.; Relloe P.; Davidson B.E.;
RT "Two promoters control the aroH gene of Escherichia coli.";
RL Gene 102:87-91 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MGL1655;
MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
RA Riley M.K.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;
RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
RA Mau B.; Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
MEDLINE=97251357; PubMed=9097039;
RA Alba H.; Baba T.; Fujita K.; Hayashi K.; Inada T.; Isono K.;
RA Itoh T.; Kasai H.; Kashimoto K.; Kimura S.; Kitakawa M.;
RA Kitagawa M.; Makino K.; Miki T.; Mizobuchi K.; Mori H.; Mori T.;
RA Motomura K.; Nakade S.; Nakamura Y.; Nishimoto H.; Nishio Y.;
RA Oshima T.; Saito N.; Sampei G.; Seki Y.; Sivasubram S.;
RA Tagami H.; Takeda J.; Takeuchi Y.; Wada C.;
RA Yamamoto Y.; Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
MEDLINE=21074935; PubMed=11206551;
RA Perna N.T.; Plunkett G. III; Burland V.; Mau B.; Glasner J.D.;
RA Rose D.J.; Mayhew G.F.; Evans P.S.; Gregor J.; Kirkpatrick H.A.;
RA Posfai G.; Hackett J.; Klink S.; Boutin A.; Shao Y.; Miller L.;
RA Grobeck E.J.; Davis N.W.; Lim A.; Dimalanta E.T.; Potamoulis K.;
RA Apodaca J.; Anantharaman T.S.; Lin J.; Yen G.; Schwartz D.C.;
RA Welch R.A.; Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509552;
MEDLINE=21156231; PubMed=11258796;
RA Hayashi T.; Makino K.; Ohnishi M.; Kurokawa K.; Ishii K.; Yokoyama K.;
RA Han C.-G.; Ohtsubo E.; Nakayama K.; Murata T.; Tanaka M.; Tobe T.;
RA Iida T.; Takami H.; Honda T.; Sasakawa C.; Ogasawara N.; Yasunaga T.;
RA Kuhara S.; Shiba T.; Hattori M.; Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M.; Rudd K.E.; Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
bacterial genome.";
RL Nucleic Acids Res. 22:4756-4767 (1994).
CC -!- SIMILARITY: TO Y. ENTEROCOLITICA HEMP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M3266; -; NOT ANNOTATED CDS.  
DR EMBL; AE000266; AAC74775.1; -;  
DR EMBL; D90813; BAA15474.1; -;  
DR EMBL; AE005394; AAG56692.1; -;  
DR EMBL; AF002558; BAB35835.1; -;  
DR FIR; A64929; A64929.  
DR FIR; D90930; D90930.  
DR FIR; H85778; H85778.  
DR EcoGene; EG12391; ydiE.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 63 AA; 7117 MW; ALEA3DD6F1DC69A1 CRC64;  
  
Query Match 61.5%; Score 24; DB 1; Length 63;  
Best Local Similarity 60.0%; Pred. No. 1.8e+02;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LEHDG 5  
Db 42 IDHDG 46  
:::|  
  
RESULT 29  
Y082\_BACAN  
ID Y082\_BACAN STANDARD; PRT; 69 AA.  
AC Q9RMV3.  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein pX02-82.  
GN PX02-82.  
OS Bacillus anthracis.  
OG Bacillus anthracis.  
OG Plasmid pX02.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1392;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,  
RA Kumanov S., Manter D., Martinez Y., Svensson R., Tatum L.R.,  
RA Brown A.E., Jackson P.J.,  
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF188935; AAF13686.1; -;  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 69 AA; 8041 MW; B69DECA31451EAF4 CRC64;  
  
Query Match 61.5%; Score 24; DB 1; Length 69;  
Best Local Similarity 66.7%; Pred. No. 2e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY -1 LEHDGI 6  
Db 50 VEHDGI 55  
:|:|:|  
  
RESULT 30  
LGEN\_BOVIN  
ID LGEN\_BOVIN STANDARD; PRT; 73 AA.  
AC P59761;

DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lactogenin (EC 3.1.27.-) (Fragments).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE, AND CHARACTERIZATION.  
RC TISSUE=Milk;  
RX MEDLINE=99417543; PubMed=10486275;  
RA Ye X.Y., Cheng K.J., Ng T.B.;  
RT "Isolation and characterization of angiogenin-1 and a novel protein  
RT designated lactogenin from bovine milk."  
RL Biochem. Biophys. Res. Commun. 263:187-191 (1999).  
RN [2]  
RP INHIBITION OF HIV-1 REVERSE TRANSCRIPTASE.  
RX MEDLINE=2055576; PubMed=11105990;  
RA Wang H., Ye X.Y., Ng T.B.;  
RT "First demonstration of an inhibitory activity of milk proteins  
RT against human immunodeficiency virus-1 reverse transcriptase and the  
RT effect of succinylation";  
RL Life Sci. 67:2745-2752 (2000).  
CC -1- FUNCTION: Secretory RNase specific towards pyrimidine bases, with  
CC higher activity towards poly C than poly U. Inhibits cell-free  
CC translation. Inhibits HIV-1 reverse transcriptase.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Milk.  
CC -1- SIMILARITY: Belongs to the pancreatic ribonuclease family.  
DR InterPro; IPR001427; RNaseA.  
DR PRODOM; PD000535; RNaseA; 1.  
DR PROSITE; PS00127; RNASE PANCREATIC; PARTIAL.  
KW Hydrolase; Nuclease; Endonuclease; Antiviral;  
KW Pyroliadone carboxylic acid.  
FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT UNSURE 36 36 OR R.  
FT UNSURE 37 37 OR E.  
FT UNSURE 38 38 OR Q.  
FT UNSURE 39 39 OR C.  
FT UNSURE 40 40 OR R.  
FT UNSURE 41 41 OR N.  
FT UNSURE 42 42 OR C.  
FT UNSURE 43 43 OR N.  
FT NON\_CONS 57 58  
FT NON\_TER 73 73  
FT SEQUENCE 73 AA; 8669 MW; 28DA862F43A9710F CRC64;  
  
Query Match 61.5%; Score 24; DB 1; Length 73;  
Best Local Similarity 60.0%; Pred. No. 2.1e+02;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 EHDGI 6  
Db 59 EHEGV 63  
||:|  
  
RESULT 31  
Z123\_HUMAN  
ID Z123\_HUMAN STANDARD; PRT; 74 AA.  
AC F35273;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc finger protein 123 (H2F-1) (Fragment).  
GN ZNF123.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=93122808; PubMed=1339395;  
RA Saleh M., Sellieri L., Little P.F., Evans G.A.;  
RT isolation and expression of linked zinc finger gene clusters on  
RT human chromosome 11q;  
RL Genomics 14:970-978(1992).  
CC -!- FUNCTION: May act as a transcription regulator in developmental  
CC processes.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- TISSUE SPECIFICITY: Liver.  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; S52506; AAB24880.1; -.  
DR PIR; A44366; A44366.  
DR Genew; HGNC:12906; ZNF123.  
DR MIM; 194630; -.  
DR GO; GO:0005634; C:nucleus; NAS.  
DR GO; GO:0003700; F:transcription factor activity; NAS.  
DR GO; GO:0008355; F:regulation of transcription, DNA-dependent; NAS.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf\_C2H2; 1.  
DR SMART; SM00355; Znf\_C2H2; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 1.

DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;  
KW Nuclear protein; Repeat.  
FT NON TER 1 1  
FT ZN FING 6 28 C2H2-TYPE.  
FT ZN FING 34 60 C2H2-TYPE.  
FT NON TER 74 74  
SQ SEQUENCE 74 AA; 8382 MW; DF3BED9F83E86FD6 CRC64;  
Query Match 61.5%; Score 24; DB 1; Length 74;  
Best Local Similarity 28.6%; Pred. No. 2.1e+02;  
Matches 2; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LEHGDGIN 7  
Db 22 IQEGVH 28  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; S52506; AAB24880.1; -.  
DR PIR; A44366; A44366.  
DR Genew; HGNC:12906; ZNF123.  
DR MIM; 194630; -.  
DR GO; GO:0005634; C:nucleus; NAS.  
DR GO; GO:0003700; F:transcription factor activity; NAS.  
DR GO; GO:0008355; F:regulation of transcription, DNA-dependent; NAS.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf\_C2H2; 1.  
DR SMART; SM00355; Znf\_C2H2; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 1.

DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;  
KW Nuclear protein; Repeat.  
FT NON TER 1 1  
FT ZN FING 6 28 C2H2-TYPE.  
FT ZN FING 34 60 C2H2-TYPE.  
FT NON TER 74 74  
SQ SEQUENCE 74 AA; 8382 MW; DF3BED9F83E86FD6 CRC64;  
Query Match 61.5%; Score 24; DB 1; Length 74;  
Best Local Similarity 28.6%; Pred. No. 2.1e+02;  
Matches 2; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AJ297529; CAC03730.1; -.  
DR HAMAP; MF 00460; -; 1.  
DR InterPro; IPR005346; UPF0125.  
DR Pfam; PF03658; UPF0125; 1.  
FT DOMAIN 79 86 POLY-ASP.  
SQ SEQUENCE 86 AA; 9688 MW; E1E3D6BED15CCD2B CRC64;

Query Match 61.5%; Score 24; DB 1; Length 86;  
Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LEHDDGI 6  
Db 30 IEHSGV 35  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U32836; AAC23277.1; -.  
DR PIR; B64133; B64133.  
DR HSSP; P02426; 1DFU.  
DR TIGR; H11630; -.  
DR InterPro; IPR001021; Ribosomal\_L25.  
DR Pfam; PF01386; Ribosomal\_L25p\_1.  
DR Pfam; PF012503; Ribosomal\_L25; 1.  
DR ProDom; PD012503; Ribosomal\_L25; 1.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 95 AA; 10652 MW; A6711F78691B2A57 CRC64;  
Query Match 61.5%; Score 24; DB 1; Length 95;

CC -!- FUNCTION: BINDS TO THE 5S RNA (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE L25P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U32836; AAC23277.1; -.  
DR PIR; B64133; B64133.  
DR HSSP; P02426; 1DFU.  
DR TIGR; H11630; -.  
DR InterPro; IPR001021; Ribosomal\_L25.  
DR Pfam; PF01386; Ribosomal\_L25p\_1.  
DR Pfam; PF012503; Ribosomal\_L25; 1.  
DR ProDom; PD012503; Ribosomal\_L25; 1.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 95 AA; 10652 MW; A6711F78691B2A57 CRC64;  
Query Match 61.5%; Score 24; DB 1; Length 95;

CC -!- FUNCTION: BINDS TO THE 5S RNA (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE L25P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
Best Local Similarity 57.1%; Pred. No. 2.8e+02; Indels 0; Gaps 0;
Matches 4; Conservative 1; Mismatches 2;

QY 1 LEHGDGIN 7
DB 43 LNHDELN 49

RESULT 34
RS6_CLOAB STANDARD; PRT; 95 AA.
ID AC Q97CX2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3US ribosomal protein S6.
GN RPSF OR CAC3724.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neelline J., Breton G., Onelchenko M.V., Makarova K.S., Zeng Q.,
Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
Tatusev R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: Binds together with S18 to 16S ribosomal RNA (By
similarity).
CC -!- SIMILARITY: Belongs to the S6P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF007867; AAKB1844.1; -
DR PIR; A97357; A97357.
DR HAMAP; MF 00360; -; 1.
DR InterPro; IPR000529; Ribosomal_S6.
DR Pfam; PF01250; Ribosomal_S6; 1.
DR ProDom; PD003809; Ribosomal_S6; 1.
DR TIGRFAMs; TIGR00166; S6; 1.
DR PROSITE; PS01048; RIBOSOMAL_S6; FALSE_NEG.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 95 AA; 10912 MW; 578189E9D9D9B08 CRC64;

Query Match 61.5%; Score 24; DB 1; Length 95;
Best Local Similarity 57.1%; Pred. No. 2.8e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHGDGIN 7
DB 14 LDEEGIN 20

RESULT 35
CMGC_BACLI STANDARD; PRT; 97 AA.
ID AC Q8QV71;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Comg operon protein 3 homolog precursor (late competence protein
comGC).
```

```
COMGC.
Bacillus licheniformis.
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14580;
RX MEDLINE=22003725; PubMed=12007649;
RA Lapidus A., Galleron N., Andersen J.T., Joergensen P.L., Ehrlich S.D.,
Sorokin A.;
RT "Co-linear scaffold of the Bacillus licheniformis and Bacillus
subtilis genomes and its use to compare their competence genes.";
RL FEMS Microbiol. Lett. 209:23-30(2002).
CC -!- FUNCTION: Required for transformation and DNA binding (By
similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: The unprocessed form is an integral membrane
protein with its C-terminus outside the membrane. Upon cleavage,
it is translocated to the outer face of the membrane (By
similarity).
CC -!- SIMILARITY: Belongs to the comGC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF459917; AAL67531.1; -
DR InterPro; IPR000983; Bac_GSPG.
DR InterPro; IPR001120; ProK_N_methyl_S.
DR PRINTS; PR00813; BCTERIALGSPG.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Competence; Transport; Methylation; Transmembrane.
FT PROPEP 1 5
FT CHAIN 6 97 COMG OPERON PROTEIN 3 HOMOLOG.
FT MOD RES 6 26 POTENTIAL.
FT MOD RES 6 6 METHYLATION (BY SIMILARITY).
FT DISULFID 41 81 BY SIMILARITY.
SQ SEQUENCE 97 AA; 10726 MW; E4B7B11257BFF3A7 CRC64;

Query Match 61.5%; Score 24; DB 1; Length 97;
Best Local Similarity 60.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHGDG 5
DB 57 MDHGDG 61

RESULT 36
PYY_DICLA STANDARD; PRT; 99 AA.
ID AC Q9FT99;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptide YY-like precursor (PYY).
GN PYY.
OS Dicertrarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Moronidae; Dicertrarchus.
OX NCBI_TaxID=13489;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98292845; PubMed=9629200;
RA Cerda-Reverter J.M., Martinez Rodriguez G., Zanuy S., Carrillo M.,
Larhammar D.;
```



RT "Cloning of neuropeptide Y, peptide YY, and peptide Y from sea bass  
RT (Dicentrarchus labrax), a marine teleost.";  
RL Ann. N.Y. Acad. Sci. 839:493-495 (1998).  
CC -|- FUNCTION: Gastrointestinal hormone and neuropeptide.  
CC -|- SUBCELLULAR LOCATION: Secreted.  
CC -|- SIMILARITY: Belongs to the NPY family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC  
CC EMBL; AJ053379; CAB4933.1; -;  
DR HSP; P01303; IRON.  
DR InterPro; IPR001955; Pancreatic\_hormn.  
DR Pfam; PF00159; hormone3; 1.  
DR PRINTS; PR00278; PANCHORMONE.  
DR ProDom; PD001267; Pancreatic\_hormn; 1.  
DR SMART; SM00309; PAH; 1.  
DR PROSITE; PS00265; PANCREATIC\_HORMONE\_1; 1.  
DR PROSITE; PS50276; PANCREATIC\_HORMONE\_2; 1.  
KW Hormone; cleavage on pair of basic residues; Amidation; Signal;  
KW Neuropeptide.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 63 PEPTIDE YY-LIKE.  
FT PROPEP 65 99 C-TERMINAL EXTENSION.  
FT MOD\_RES 63 63 AMIDATION (G-64 PROVIDE AMIDE GROUP)  
FT (POTENTIAL).  
SQ SEQUENCE 99 AA; 11065 MW; 6897DEBABD4E74FD CRC64;  
Query Match 61.5%; Score 24; DB 1; Length 99;  
Best Local Similarity 80.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 HDGIN 7  
Db 22 HSGIN 26  
RESULT 37  
FER2\_CAUCR STANDARD; PRT; 106 AA.  
AC P37058;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ferredoxin, 2Fe-2S (FdII).  
GN FDXB OR FDXP OR C3524.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]\_TaxID=155892;  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RA Amemiya K.;  
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95270612; PubMed=77511304;  
RA Wang S.P., Chen Y.P., Ely B.;  
RT "A ferredoxin, designated FdxP, stimulates p-hydroxybenzoate  
RT hydroxylase activity in Caulobacter crescentus.";  
RL J. Bacteriol. 177:2908-2911 (1995).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Uitterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).  
CC -|- COFACTOR: Binds 1 2Fe-2S cluster (By similarity).  
CC -|- SIMILARITY: Belongs to the adrenodoxin / putidaredoxin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC  
CC EMBL; X51607; CA35950.1; -;  
DR EMBL; AE006011; AAK25486.1; -;  
DR PIR; S32573; S32573.  
DR HSP; P00259; LGPX.  
DR TIGR; CG3524; -;  
DR InterPro; IPR001055; Adrenodoxin.  
DR InterPro; IPR001041; Ferredoxin.  
DR Pfam; PF00111; fer2; 1.  
DR PROSITE; PS00355; ADRENODOXIN.  
DR PROSITE; PS00814; ADX; 1.  
KW Electron transport; Metal-binding; Iron-sulfur; Iron; 2Fe-2S;  
KW Complete proteome.  
FT METAL 40 40 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 46 46 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 49 49 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 86 86 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
SQ SEQUENCE 106 AA; 11366 MW; 59D096A72D993339 CRC64;  
Query Match 61.5%; Score 24; DB 1; Length 106;  
Best Local Similarity 60.0%; Pred. No. 3.2e+02;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 LEHDG 5  
Db 7 IQHGD 11  
RESULT 38  
GCAD\_BACCL STANDARD; PRT; 116 AA.  
AC P42817;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (N-  
DE acetylglucosamine-1-phosphate uridylyltransferase) (Fragment).  
GN GCAD.  
OS Bacillus caldolyticus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 405;  
RX MEDLINE=97075912; PubMed=8918235;  
RA Krath B.N., Hove-Jensen B.;  
RT "Bacillus caldolyticus pfs gene encoding phosphoribosyl-diphosphate  
RT synthase.";  
RL Gene 176:73-79 (1996).  
CC -|- CATALYTIC ACTIVITY: UTP + N-acetyl-alpha-D-glucosamine 1-phosphate  
CC = diphosphate + UDP-N-acetyl-D-glucosamine.  
CC -|- PATHWAY: Peptidoglycan and lipopolysaccharide biosynthesis.  
CC -|- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NOCL FAMILY OF  
CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X83708; CAA58681.1; -  
 DR PIR; PC4228; PC4228.  
 DR InterPro; IPR001451; Hexapep transf.  
 DR PROSITE; PS00101; HEXAPEP TRANSFERASES; 1.  
 DR Peptidoglycan synthesis; Cell wall; Transferase;  
 KW Nucleotidyltransferase; Repeat.  
 FT NON TER 1  
 SQ SEQUENCE 116 AA; 12106 MW; D27706B9B8F7B362 CRC64;  
 -----  
 Query Match 61.5%; Score 24; DB 1; Length 116;  
 Best Local Similarity 42.9%; Pred. No. 3.5e+02;  
 Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 -----  
 QY 1 LEHDGIN 7  
 Db 42 VNYDGVN 48  
 : : : : :  
 : : : : :  
 -----  
 RESULT 39  
 FOLD STRCO STANDARD; PRT; 119 AA.  
 ID FOLB STROCO  
 AC QX8IO;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Dihydroneopterin aldolase (EC 4.1.2.25) (DHNA).  
 GN FOLB OR SCO3400 OR SCE9.07.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=A3(2) / M145;  
 RC MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Gobie A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -!- FUNCTION: Catalyzes the conversion of 7,8-dihydroneopterin to 6-  
 CC hydroxymethyl-7,8-dihydropterin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: 2-amino-4-hydroxy-6-(D-erythro-1,2,3-  
 CC trihydroxypropyl)-7,8-dihydropteridine = 2-amino-4-hydroxy-6-  
 CC hydroxymethyl-7,8-dihydropteridine + glycolaldehyde.  
 CC -!- PATHWAY: Folate biosynthesis.  
 CC -!- SIMILARITY: Belongs to the DHNA family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL939116; CAB42753.1; -  
 DR PIR; T36326; T36326.  
 DR HSSP; P56740; 1DHN.

DR InterPro; IPR006157; FOLB.  
 DR InterPro; IPR006156; FOLB\_fam.  
 DR Pfam; PF02152; FOLB; 1.  
 DR TIGRFAMs; TIGR00525; folB; 1.  
 DR TIGRFAMs; TIGR00526; folB\_dom; 1.  
 KW Lyase; Folate biosynthesis; Complete proteome.  
 SQ SEQUENCE 119 AA; 13002 MW; 0E86573FA1A694D CRC64;  
 -----  
 Query Match 61.5%; Score 24; DB 1; Length 119;  
 Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 1 LEHDGI 6  
 Db 84 LKHEGV 89  
 : : : : :  
 : : : : :  
 -----  
 RESULT 40  
 PHNA HAEIN STANDARD; PRT; 120 AA.  
 ID PHNA HAEIN  
 AC P44479;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PhnA protein homolog.  
 DE PHNA OR H10046.  
 GN Haemophilus influenzae.  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=rd / KW20 / ATCC 51907;  
 RC MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.D., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32690; AAC21724.1; -  
 DR PIR; I64044; I64044.  
 DR TIGR; H10046; -  
 DR InterPro; IPR004624; PhnA.  
 DR Pfam; PF03831; PhnA; 1.  
 DR TIGRFAMs; TIGR00686; phnA; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 120 AA; 13678 MW; 5553FE2ACC6C6245 CRC64;  
 -----  
 Query Match 61.5%; Score 24; DB 1; Length 120;  
 Best Local Similarity 80.0%; Pred. No. 3.7e+02;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 -----  
 QY 3 HDGIN 7  
 Db 29 HDSIN 33  
 : : : : :  
 : : : : :  
 -----

```

RESULT 41
YDEP_BACSU STANDARD; PRT; 128 AA.
ID YDEP_BACSU STANDARD; PRT; 128 AA.
AC P96673; C31495;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ydeP.
GN YDEP OR BSU05290.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kasahara Y., Nakai S., Lee S., Sadaie Y., Ogasawara N.;
RT "A 148 kbp sequence of the region between 35 and 47 degree of the
RL Bacillus subtilis genome.";
RN Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerthon I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrazi E., Foulger D.,
RA Friiz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koeter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maul C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudea S., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: Belongs to the UPF0087 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB001489; BAI19363.1; ALT_INIT.
DR EMBL; Z99106; CAB12336.1; -.
DR PIR; B69779; B69779.
DR Subtilist; BG12143; ydeP.
DR InterPro; IPR002577; DUF24.
DR Pfam; PF01638; DUF24; 1.
DR ProDom; PD004032; DUF24; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 128 AA; 15211 MW; A47D416545CCA5C8 CRC64;

```

```

Query Match 61.5%; Score 24; DB 1; Length 128;
Best Local Similarity 66.7%; Pred. No. 3.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDGI 6
Db 68 LERDGV 73

RESULT 42
MRKF KLEPN STANDARD; PRT; 132 AA.
ID MRKF KLEPN STANDARD; PRT; 132 AA.
AC P21650;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE MrKF protein.
GN MRKF.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IA565;
RX MEDLINE=91100388; PubMed=1670938;
RA Allen B.L., Gerlach G.-F., Clegg S.;
RT "Nucleotide sequence and functions of mrk determinants necessary for
RL expression of type 3 fimbriae in Klebsiella pneumoniae.";
RL J. Bacteriol. 173:916-920(1991).
CC -!- FUNCTION: APPEARS TO AFFECT THE STABILITY OF THE INTACT FIMBRIAE
CC ON THE CELL SURFACE.
CC -!- SUBCELLULAR LOCATION: Fimbria.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M55912; AAA25097.1; -.
DR PIR; F39142; F39142.
DR InterPro; IPR008966; Adhes_bact.
KW Fimbria.
SQ SEQUENCE 132 AA; 14531 MW; 1C821B681912BC1A CRC64;

Query Match 61.5%; Score 24; DB 1; Length 132;
Best Local Similarity 80.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDG 5
Db 15 LRHDG 19

RESULT 43
RS12 ERYGR STANDARD; PRT; 132 AA.
ID RS12 ERYGR STANDARD; PRT; 132 AA.
AC OS9936;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 40S ribosomal protein S12.
GN RPS12 OR ERPR1.
OS Erysiphe graminis (subsp. hordei) (Grass mildew) (Blumeria graminis
OC (subsp. hordei)).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
OC Erysiphales; Erysiphaceae; Blumeria.
OX NCBI_TaxID=62688;
RN [1]

```

RP SEQUENCE FROM N.A.  
RA Zhang Z., Gurr S.J.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 12-132 FROM N.A.  
RA Zhang Z.G., Gurr S.J.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: Belongs to the S12E family of ribosomal proteins.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF052483; AAC15834.1; -;  
DR EMBL; AF051259; AAC15802.1; -;  
DR InterPro; IPR004038; Ribosomal L7A.  
DR Pfam; PF01248; Ribosomal L7Ae; 1.  
DR PRINTS; PR00972; RIBOSOMAL\_S12E.  
DR PROSITE; PS01189; RIBOSOMAL\_S12E; 1.  
KW Ribosomal protein.  
SQ SEQUENCE 132 AA; 14632 MW; 91F3F48DA07917B4 CRC64;  
  
Query Match 61.5%; Score 24; DB 1; Length 132;  
Best Local Similarity 66.7%; Pred. No. 4.1e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LEHGI 6  
DB 26 LMHDGL 31  
  
RESULT 44  
SPAK\_SHIFL STANDARD; PRT; 133 AA.  
AC P35530;  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Surface presentation of antigens protein spak (Spa15 protein).  
GN SPK OR SPA15 OR CP0148.  
OS Shigella flexneri.  
OG Plasmid pWR100, Plasmid pMYSH6000, and Plasmid pCP301.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M90T / Serotype 5; PLASMID=pWR100;  
RX MEDLINE=92193289; PubMed=1312536;  
RA Venkatesan M.M., Buysse J.M., Oaks E.V.;  
RT "Surface presentation of Shigella flexneri invasion plasmid antigens  
RT requires the products of the spa locus."  
RL J. Bacteriol. 174:1990-2001(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M90T / Serotype 5; PLASMID=pWR100;  
RX MEDLINE=20566792; PubMed=1115111;  
RA Buchrieser C., Glaser P., Rusniok C., Nedjari H., d'Hauteville H.,  
RA Kunst F., Sansonetti P., Parsot C.;  
RT "The virulence plasmid pWR100 and the repertoire of proteins secreted  
RT by the type III secretion apparatus of Shigella flexneri";  
RL Mol. Microbiol. 38:760-771(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M90T / Serotype 5; PLASMID=pWR100;  
RX MEDLINE=21189246; PubMed=11292750;  
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,  
RA Blattner F.R.;

RT "Complete DNA sequence and analysis of the large virulence plasmid of  
RT Shigella flexneri".  
RL Infect. Immun. 69:3271-3285(2001).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YSH6000 / Serotype 2a; PLASMID=pMYSH6000;  
RX MEDLINE=93224456; PubMed=8385666;  
RA Sasaki K., Komatsu K., Tobe T., Suzuki T., Yoshikawa M.;  
RT "Eight genes in region 5 that form an operon are essential for  
RT invasion of epithelial cells by Shigella flexneri 2a".  
RL J. Bacteriol. 175:2334-2346(1993).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=301 / Serotype 2a; PLASMID=pCP301;  
RX MEDLINE=2272406; PubMed=12384590;  
RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
RT through comparison with genomes of Escherichia coli K12 and O157".  
RL Nucleic Acids Res. 30:4432-4441(2002).  
CC -!- FUNCTION: REQUIRED FOR SURFACE PRESENTATION OF INVASION PLASMID  
CC ANTIGENS. COULD PLAY A ROLE IN PRESERVING THE TRANSLOCATION  
CC COMPETENCE OF THE IPA ANTIGENS.  
CC -!- SIMILARITY: BELONGS TO THE SPK FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M81458; AAA26540.1; -;  
DR EMBL; AL391753; CAC05823.1; -;  
DR EMBL; AF348706; AAK18467.1; -;  
DR EMBL; D13663; BAA02824.1; -;  
DR EMBL; AF386526; AAL72308.1; -;  
DR PIR; B42284; B42284.  
DR InterPro; IPR003065; Invas Spak.  
DR Pfam; PF03519; Invas Spak; 1.  
DR PRINTS; PR01305; SSPAKPROTEIN.  
DR ProDom; PD016047; Invas Spak; 1.  
KW Plasmid; Virulence; Complete proteome.  
SQ SEQUENCE 133 AA; 15116 MW; EA8F145A98A002B1 CRC64;  
  
Query Match 61.5%; Score 24; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 HDGI 6  
DB 110 HDGI 113  
  
RESULT 45  
SPAK\_SHISO STANDARD; PRT; 133 AA.  
AC Q55296;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Surface presentation of antigens protein spak (Spa15 protein).  
GN SPK OR SPA15.  
OS Shigella sonnei.  
OG Plasmid.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=624;  
RN [1]

RP SEQUENCE FROM N.A.  
RA Arakawa E., Kato J.I., Ito K.I., Watanabe H.;  
RT "Comparison and high conservation of nucleotide sequences of spa-mxi  
RT regions between S.sonnei and S.flexneri -- identification of a new  
RT gene coding plausible membrane protein."  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR SURFACE PRESENTATION OF INVASION PLASMID  
CC ANTIGENS. COULD PLAY A ROLE IN PRESERVING THE TRANSLOCATION  
CC COMPETENCE OF THE IPA ANTIGENS.  
CC -!- SIMILARITY: BELONGS TO THE SPAK FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D50601; BAA09157.1; -  
DR InterPro; IPR003065; Invas\_Spak.  
DR Pfam; PF03519; Invas\_Spak; 1.  
DR PRINTS; PRO1305; SSPAKPROTEIN.  
DR ProDom; PD016047; Invas\_Spak; 1.  
KW Plasmid; Virulence.  
SQ SEQUENCE 133 AA; 15188 MW; EA9F173A85A002B1 CRC64;  
  
Query Match 61.5%; Score 24; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 HDGI 6  
Db 110 HDGI 113

Search completed: June 15, 2004, 12:23:19  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2004, 12:22:52 ; Search time 40 Seconds  
(without alignments)  
55.216 Million cell u

Title: US-09-394-019C-212  
Perfect score: 39  
Sequence: 1 LEHDGIN 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 311989

```
Minimum DB seq length: 0
Maximum DB seq length: 150
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 s

Database : SPTREMBL 25: \*

- ```

1:  sp_archea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query No. | Score | Query |        |        | ID                   | Description |
|------------|-----------|-------|-------|--------|--------|----------------------|-------------|
|            |           |       | Match | Length | DB     |                      |             |
| 1          | 32        | 82.1  | 128   | 17     | Q9YAM4 | Q9Yam4 aeropyrum p   |             |
| 2          | 31        | 79.5  | 67    | 17     | Q8UIM7 | Q8uim7 pyrococcus    |             |
| 3          | 30        | 76.9  | 63    | 10     | Q9PTE4 | Q9fte4 oryza sativ   |             |
| 4          | 30        | 76.9  | 64    | 11     | Q8YI59 | Q8bct59 mus musculus |             |
| 5          | 30        | 76.9  | 78    | 16     | Q8XSB6 | Q8xsb6 ralstonia s   |             |
| 6          | 30        | 76.9  | 80    | 16     | Q9KRZ6 | Q9krz6 vibrio chol   |             |
| 7          | 30        | 76.9  | 80    | 16     | Q8PBA1 | Q8pba1 xanthomonas   |             |
| 8          | 30        | 76.9  | 80    | 16     | Q8UWJ1 | Q8uwj1 gallus gall   |             |
| 9          | 30        | 76.9  | 97    | 16     | Q9L2K5 | Q9l2k5 streptomyce   |             |
| 10         | 30        | 76.9  | 114   | 16     | Q8D0R6 | Q8dgr6 streptococc   |             |
| 11         | 30        | 76.9  | 122   | 12     | Q3J5B5 | Q9j5es fowlbox vir   |             |
| 12         | 30        | 76.9  | 136   | 10     | Q8LQ98 | Q8lq98 oryza sativ   |             |
| 13         | 30        | 76.9  | 142   | 12     | Q91BT7 | Q91bt7 helicoverpa   |             |
| 14         | 30        | 76.9  | 144   | 12     | Q9E2I5 | Q9e2i5 helicoverpa   |             |
| 15         | 30        | 76.9  | 144   | 12     | Q91FI9 | Q91fi9 helicoverpa   |             |
| 16         | 30        | 76.9  | 146   | 16     | Q8CWE3 | Q8cwe3 escherichia   |             |

|    |    |      |     |    |        |                     |
|----|----|------|-----|----|--------|---------------------|
| 17 | 29 | 74.4 | 55  | 11 | Q99JC7 | Q99JC7 rattus norv  |
| 18 | 29 | 74.4 | 59  | 6  | Q95KU8 | Q95KU8 bos taurus   |
| 19 | 29 | 74.4 | 67  | 2  | Q47667 | Q47667 escherichia  |
| 20 | 29 | 74.4 | 68  | 13 | Q42337 | Q42337 xenopus lae  |
| 21 | 29 | 74.4 | 73  | 2  | Q936F7 | Q936F7 staphylococ  |
| 22 | 29 | 74.4 | 77  | 16 | Q3KCM0 | Q3KCM0 chlorobium   |
| 23 | 29 | 74.4 | 82  | 16 | Q38W3  | Q38W3 enterococcu   |
| 24 | 29 | 74.4 | 88  | 7  | Q860C7 | Q860C7 acrocephalu  |
| 25 | 29 | 74.4 | 88  | 7  | Q860C6 | Q860C6 acrocephalu  |
| 26 | 29 | 74.4 | 88  | 7  | Q860C5 | Q860C5 acrocephalu  |
| 27 | 29 | 74.4 | 88  | 7  | Q860C4 | Q860C4 acrocephalu  |
| 28 | 29 | 74.4 | 88  | 7  | Q860C3 | Q860C3 acrocephalu  |
| 29 | 29 | 74.4 | 90  | 7  | Q860C9 | Q860C9 acrocephalu  |
| 30 | 29 | 74.4 | 97  | 6  | Q8SPLO | Q8SPLO equus caball |
| 31 | 29 | 74.4 | 97  | 6  | Q8SPJ5 | Q8SPJ5 sus scrofa   |
| 32 | 29 | 74.4 | 98  | 6  | Q9N1C0 | Q9N1C0 ovis aries   |
| 33 | 29 | 74.4 | 100 | 4  | Q9UMB0 | Q9UMB0 homo sapien  |
| 34 | 29 | 74.4 | 109 | 16 | Q9PIF8 | Q9PIF8 xanthomonas  |
| 35 | 29 | 74.4 | 117 | 9  | Q8W6N9 | Q8W6N9 bacterioph   |
| 36 | 29 | 74.4 | 117 | 16 | Q9PE20 | Q9PE20 xanthomonas  |
| 37 | 29 | 74.4 | 124 | 2  | Q8XSU0 | Q8XSU0 rhizobium l  |
| 38 | 29 | 74.4 | 137 | 4  | Q9H8N2 | Q9H8N2 homo sapien  |
| 39 | 29 | 74.4 | 138 | 12 | Q9DHI4 | Q9DHI4 yaba-like d  |
| 40 | 29 | 74.4 | 141 | 5  | Q9V1T3 | Q9V1T3 jasus edwar  |
| 41 | 29 | 74.4 | 146 | 17 | Q97XH7 | Q97XH7 sulfobobus   |
| 42 | 29 | 74.4 | 146 | 17 | Q972G0 | Q972G0 sulfobobus   |
| 43 | 29 | 74.4 | 147 | 16 | Q89FE7 | Q89FE7 bradyrhizob  |
| 44 | 28 | 71.8 | 22  | 2  | Q34195 | Q34195 ehrlichia r  |
| 45 | 28 | 71.8 | 69  | 16 | Q81RR9 | Q81RR9 bacillus an  |

## ALIGNMENTS

## RESULT 1

Q9YAM4  
ID Q9YAM4  
PRELIMINARY;  
PRT; 128 AA.

|         |                                                            |                                         |
|---------|------------------------------------------------------------|-----------------------------------------|
| Q91AW4? | 01-NOV-1999                                                | (TrEMBLrel. 12, Created)                |
| DT      | 01-NOV-1999                                                | (TrEMBLrel. 12, Last sequence update)   |
| DT      | 01-JUN-2003                                                | (TrEMBLrel. 24, Last annotation update) |
| DT      | 01-JUN-2003                                                | (TrEMBLrel. 24, Last annotation update) |
| DE      | Hypothetical protein APE1919.                              |                                         |
| DN      | APE1919.                                                   |                                         |
| GE      |                                                            |                                         |
| OS      | Aeropyrum pernix.                                          |                                         |
| OC      | Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;   |                                         |
| OC      | Desulfurococcales; Aeropyrum.                              |                                         |
| OX      | NCBI_Taxid=56636;                                          |                                         |
| OX      | [1]                                                        | RN                                      |
| PC      | SEQUENCE FROM N.A.                                         |                                         |
| RP      | STRAIN=K1.                                                 |                                         |
| RC      |                                                            |                                         |
| RX      | MEDLINE=99a10339;                                          | PubMed=10382966;                        |
| RA      | Kawarayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa |                                         |
| RA      | Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A.,  |                                         |
| RA      | Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H. |                                         |
| RA      | Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y., |                                         |
| RA      | Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K., |                                         |
| RA      | Nakamura Y., Nomura N., Sako Y., Kikuchi H.,               |                                         |
| RT      | "Complete genome sequence of an aerobic hyper-thermophilic |                                         |
| RT      | crenarchaeon, Aeropyrum pernix K1.";                       |                                         |
| RL      | DNA Res. 6:83-101(1999).                                   |                                         |
| RL      | ENBL; AF000062; BAA0924.1; -.                              |                                         |
| DR      | PIR; G72579; G72579.                                       |                                         |
| DR      | PIR; G72579; G72579.                                       |                                         |
| KW      | Hypothetical protein; Complete proteome.                   |                                         |
| SO      | SEQUENCE 128 AA; 13027 MW; 20FC54839DE21121; CRC64;        |                                         |

Query Match 82.1%; Score 32; DB 17; Length 128;  
Best Local Similarity 71.4%; Pred. No. 72;  
Matches 5. Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LEHDGIN 7  
116 LSHDGLN 122  
Db

```

RESULT 2
Q8UIM7
ID Q8UIM7 PRELIMINARY; PRT; 67 AA.
AC Q8UIM7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PF1179.
GN PF1179.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 9422;
RA Weiss R.B., Dunn D.M., Robb E.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010226; AAL81303.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 67 AA; 7561 MW; 591FE1D4A09AF560 CRC64;

Query Match 79.5%; Score 31; DB 17; Length 67;
Best Local Similarity 83.3%; Pred. No. 56;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHGGI 6
DB 41 LEHDDL 46

RESULT 3
Q9FTE4
ID Q9FTE4 PRELIMINARY; PRT; 63 AA.
AC Q9FTE4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P0698A04.7 protein (P0494A10.23 protein).
GN P0698A04.7 OR P0494A10.23.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0698A04.7."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0494A10.23."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002868; BAB17146.1; -.
DR EMBL; AP002541; BAB55483.1; -.
DR Gramene; Q9FTE4; -.
SQ SEQUENCE 63 AA; 6601 MW; 294AC4BDEC79948D CRC64;

Query Match 76.9%; Score 30; DB 10; Length 63;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7
DB 58 HDGIN 62

```

```

RESULT 4
Q8BT59
ID Q8BT59 PRELIMINARY; PRT; 64 AA.
AC Q8BT59;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Junction cell adhesion molecule 3 (Fragment).
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK017692; BAC25526.1; -.
FT NON TER 1
SQ SEQUENCE 64 AA; 6996 MW; AF46BC30AA6D0C11 CRC64;

Query Match 76.9%; Score 30; DB 11; Length 64;
Best Local Similarity 66.7%; Pred. No. 87;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGIN 7
DB 41 KHDGVN 46

RESULT 5
Q8XSB6
ID Q8XSB6 PRELIMINARY; PRT; 78 AA.
AC Q8XSB6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein RSP0562.
GN RSP0562 OR RS03916.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
Chandler M., Choise N., Claudel-Renard C., Cunac S., Denange N.,
Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502 (2002).
DR EMBL; AL646079; CAD17713.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR002577; DUF24.
DR Pfam; PF01638; DUF24; 1.
DR ProDom; PD004032; DUF24; 1.
DR Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 78 AA; 8918 MW; 0AE00EAB5C9A3B0A CRC64;

Query Match 76.9%; Score 30; DB 16; Length 78;
Best Local Similarity 83.3%; Pred. No. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 LEHNGI 6
Db      56 LEHNGM 61

RESULT 6
ID Q9KRZ6 PRELIMINARY; PRT; 80 AA.
AC Q9KRZ6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein VC1487.
GN VC1487.cholerae.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20405833; PubMed=10952301;
RA Heidelberg J.F., Bisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AB004226; AAF94642.1; -.
DR PIR; H82194; H82194.
DR TIGR; VC1487; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 80 AA; 9275 MW; F6B667C37C4F6892 CRC64;

Query Match 76.9%; Score 30; DB 16; Length 80;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 LEHNGIN 7
Db      71 LKHNGIN 77

RESULT 7
ID Q8PBA1 PRELIMINARY; PRT; 80 AA.
AC Q8PBA1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE HrpD6 protein
GN HRPD6 OR XC1222.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Farfa J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AB012221; AAM40520.1; -.
KW Complete proteome.
SQ SEQUENCE 80 AA; 8787 MW; E723B262314DALLE CRC64;

Query Match 76.9%; Score 30; DB 16; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 HDGIN 7
Db      63 HDGIN 67

RESULT 8
ID Q8UWJ1 PRELIMINARY; PRT; 89 AA.
AC Q8UWJ1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Bcl-x (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Shi Z., Onagbesan O.M., Williams J.;
RT "Apoptosis in chicken ovary."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF432511; AAL35559.1; -.
DR GO; GO:0016329; F:apoptosis regulator activity; IEA.
DR GO; GO:0006915; P:apoptosis; IEA.
DR InterPro; IPR00712; BCL2_BH.
DR InterPro; IPR02475; BCL2_family.
DR Pfam; PF00452; Bcl-2; 1.
DR SMART; SM00337; BCL; 1.
DR PROSITE; PS50062; BCL2_FAMILY; 1.
DR PROSITE; PS01080; BH1; 1.
DR PROSITE; PS01259; BH3; 1.
FT NON_TER 1
FT NON_TER 89
SQ SEQUENCE 89 AA; 10124 MW; B5B0EBE5F323A8C4 CRC64;

Query Match 76.9%; Score 30; DB 13; Length 89;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 LEHNGIN 7
Db      50 LFHDGIN 56

RESULT 9
ID Q9L2K5 PRELIMINARY; PRT; 97 AA.
AC Q9L2K5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SCO0699.

```



SC00699 OR SCF42.09C.  
 GN Streptomyces coelicolor.  
 OS Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939106; CAB69671.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 97 AA; 10089 MW; 4CFD924E273194F5 CRC64;

Query Match 76.9%; Score 30; DB 16; Length 97;  
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDSGI 6  
 |:|||||  
 Db 35 LKHDGI 40

RESULT 10  
 Q8DQ6 PRELIMINARY; PRT; 114 AA.  
 AC Q8DQ6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter ATP-binding protein-glutamine, truncation.  
 GN GLN-TRUNCATION OR SPR0535.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
 RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,  
 RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL; AE008432; AAK93339.1; -.  
 DR PIR; G97938; G97938.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 KW ATP-binding; Complete proteome.  
 SQ SEQUENCE 114 AA; 12665 MW; 32315A476EF6AD51 CRC64;

Query Match 76.9%; Score 30; DB 16; Length 114;  
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDSGIN 7  
 |:|||||  
 Db 47 LAHDGWN 53

RESULT 11  
 Q9J5B5 PRELIMINARY; PRT; 122 AA.  
 AC Q9J5B5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE ORF FPV066 hypothetical protein.  
 GN FPV066.  
 OS Fowlpox virus (FPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Avipoxvirus.  
 OX NCBI\_TaxID=10261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20193820; PubMed=10729156;  
 RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
 RT "The genome of fowlpox virus.";  
 RL J. Virol. 74:3815-3831(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AF198100; AAF44410.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 122 AA; 14797 MW; 908629B0DB871662 CRC64;

Query Match 76.9%; Score 30; DB 12; Length 122;  
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDSGIN 7  
 |:|||||  
 Db 10 LKHDGIS 16

RESULT 12  
 Q8LQ98 PRELIMINARY; PRT; 136 AA.  
 AC Q8LQ98;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE B1011A07.14 protein.  
 GN B1011A07.14.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC  
 clones:B1011A07.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AF003722; BAB92766.1; -.  
 DR Gramene; Q8LQ98; -.  
 SQ SEQUENCE 136 AA; 14422 MW; 07F19447102FD19D CRC64;

Query Match 76.9%; Score 30; DB 10; Length 136;  
 Best Local Similarity 57.1%; Pred. No. 2e+02;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7

DB 121 VSHDGVN 127

RESULT 13

Q91BT7

ID Q91BT7 PRELIMINARY; PRT; 142 AA.

AC Q91BT7 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Helicoverpa armigera nuclear polyhedrosis virus.

OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OC NCBI\_TaxID=51313;

OX [1]

RN SEQUENCE FROM N.A.

RP STRAIN=C1;

RC PubMed=12050807;

RX Zhang C.X., Wu J.C.;

RA "Genome structure and the p10 gene of the Helicoverpa armigera nucleopolyhedrovirus.";

RT Sheng Wu Hua Haueh Yu Sheng Wu Wu Li Hsueh Pao 33:179-184 (2001).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=C1;

RC Zhang C.X., Jin W.R.;

RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL; AF303045; AAK96369.1; -

DR Hypothetical protein.

KW Hypothetical protein.

SQ SEQUENCE 142 AA; 16036 MW; 45D3B56DAE934C82 CRC64;

Query Match 76.9%; Score 30; DB 12; Length 142;

Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7

DB 60 HDGIN 64

RESULT 14

Q9E215

ID Q9E215 PRELIMINARY; PRT; 144 AA.

AC Q9E215 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Orf435-like protein (ORF129).

OS Helicoverpa zea single nucleocapsid nucleopolyhedrovirus.

OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OC NCBI\_TaxID=10468;

OX [1]

RN SEQUENCE FROM N.A.

RP Robertson A.P.S.;

RA "Genetic organization of Helicoverpa zea nuclear polyhedrosis virus in the region of EcoRI U,D,L,A and Q.";

RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RL [2]

RN SEQUENCE FROM N.A.

RP Chen X., Zhang W.-J., Wong J., Chun G., Lu A., McCutchen B.F., Presnail J.K., Herrmann R., Dotan M., Tingey S., Hu Z.-H., Vlak J.M.;

RA "Genome sequence analysis of Helicoverpa zea single nucleocapsid nucleopolyhedrovirus.";

RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RL EMBL; AF275264; AAG17387.1; -

DR EMBL; AF334030; AAL56135.1; -

SQ SEQUENCE 144 AA; 16214 MW; A38170147F20782A CRC64;

Query Match 76.9%; Score 30; DB 12; Length 144;

Best Local Similarity 100.0%; Pred. No. 2.2e+02; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7

DB 60 HDGIN 64

RESULT 15

Q91FT9

ID Q91FT9 PRELIMINARY; PRT; 144 AA.

AC Q91FT9 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ORF435 (Hypothetical protein).

OS Helicoverpa armigera nuclear polyhedrosis virus, and

OS Helicoverpa armigera nucleopolyhedrovirus G4.

OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OC NCBI\_TaxID=51313, 148363;

OX [1]

RN SEQUENCE FROM N.A.

RP SPECIES=Helicoverpa armigera nuclear polyhedrosis virus;

RC MEDLINE=20033815; PubMed=10584750;

RA Chen X., Ijkel W.F., Dominy C., de Andrade Zanotto P.M., Hashimoto Y., Faktor O., Hayakawa T., Wang C.H., Prekumar A., Mathavan S., Krell P.J., Hu Z., Vlak J.M.;

RA "Identification, sequence analysis and phylogeny of the lef-2 gene of Helicoverpa armigera single-nucleocapsid baculovirus.";

RL Virus Res. 65:21-32(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus;

RA Long G., Chen X., Vlak J.M., Hu Z.;

RA "Genetic organization of the HindIII-I region of Helicoverpa armigera single-nucleocapsid nucleopolyhedrovirus.";

RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RL [3]

RP SEQUENCE FROM N.A.

RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;

RA Deng F., Chen X., Vlak J.M., Arif B.M., Hu Z.;

RA "Sequence analysis of the gp37 gene of Heliothis armigera single-nucleocapsid nucleopolyhedrovirus.";

RT Zhongguo Bingdaxue 15:35-42(2000).

RL [4]

RP SEQUENCE FROM N.A.

RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;

RA Wang H., Hu Z., Sun X., Vlak J.M., Chen X.;

RA "Sequence analysis of the iap3 gene of Heliothis armigera single-nucleocapsid nucleopolyhedrovirus.";

RT Zhongguo Bingdaxue 15:43-49(2000).

RL [5]

RP SEQUENCE FROM N.A.

RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;

RA MEDLINE=21078302; PubMed=11210934;

RA Wang H., Chen X., Wang H., Arif B.M., Vlak J.M., Hu Z.;

RA "Nucleotide sequence and transcriptional analysis of a putative basic DNA-binding protein of Helicoverpa armigera polyhedrovirus.";

RT Virus Genes 22:113-120(2001).

RL [6]

RN SEQUENCE FROM N.A.

RP SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;

RC PubMed=11125177;

RA Chen X., Ijkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H., Peters S., Zuidema D., Lankhorst R.K., Vlak J.M., Hu Z.;

RA "The sequence of the Helicoverpa armigera single-nucleocapsid nucleopolyhedrovirus genome.";

RT nucleopolyhedrovirus genome.";

RL J. Gen. Virol. 82:241-257(2001).

RN [7]

RP SEQUENCE FROM N.A.

RC SPECIES=Helicoverpa amigera nucleopolyhedrovirus G4;  
RA Chen X., Ijkel W.F., Turchini R., Sun X., Sandbrink H., Wang H.,  
RA Peters S., Zuidema D., Lankhorst R.K., Vlak J.M., Hu Z.,  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF266697; AAF78940.1; -;  
DR EMBL; AF271059; AAG53868.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 144 AA; 16115 MW; A3816C6476506D2A CRC64;  
  
Query Match 76.9%; Score 30; DB 12; Length 144;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 HDGIN 7  
Db 60 HDGIN 64  
  
RESULT 16  
Q8CWE3 PRELIMINARY; PRT; 146 AA.  
ID Q8CWE3  
AC Q8CWE3  
DT 01-NAR-2003 (TREMELrel. 23, Created)  
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Prepilin peptidase dependent protein D precursor.  
GN PPDB OR C0127.  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAINS=O6:H1 / CPT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA McInley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
DR EMBL; AB016755; AAN78625.1; -;  
DR GO; GO:0009289; C:filmbria; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001102; Pili.  
DR InterPro; IPR001120; Prok\_N\_methyl\_S.  
DR Pfam; PF00114; Pili; 1.  
DR ProDom; PD000666; Pili; 1.  
DR PROSITE; PS00409; PROKAR\_NTER\_METHYL; 1.  
KW Complete proteome.  
SQ SEQUENCE 146 AA; 15608 MW; 2305DABEA13F6ADCD CRC64;

Query Match 76.9%; Score 30; DB 16; Length 146;  
Best Local Similarity 71.4%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LEHGIN 7  
Db 58 LEHGIN 64

RESULT 17  
Q99JC7 PRELIMINARY; PRT; 55 AA.  
ID Q99JC7  
AC Q99JC7  
DT 01-JUN-2001 (TREMELrel. 17, Created)  
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Soluble fibroblast growth factor receptor IIb (Fragment).  
GN SKGF-R.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar; TISSUE=Testis;  
RA Konrad L.;  
RT "Expression of soluble FGF-R IIb in testicular cell lines.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ312745; CAC37408.1; -;  
DR GO; GO:0004872; F:receptor activity; IEA.  
DR InterPro; IPR007110; Ig-like.  
KW Receptor.  
KW NON\_TER  
SQ SEQUENCE 55 AA; 6083 MW; 69886E17147DC94E CRC64;  
  
Query Match 74.4%; Score 29; DB 11; Length 55;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LEHGIN 7  
Db 1 LKHGIN 7  
  
RESULT 18  
Q95KU8 PRELIMINARY; PRT; 59 AA.  
ID Q95KU8  
AC Q95KU8  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Fibroblast growth factor receptor 2 (Fragment).  
GN FGFR2.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Berisha B.A., Schams D.;  
RT "Expression of fibroblast growth factor family in the bovine placenta  
during the pregnancy.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ419173; CAD11602.1; -;  
DR GO; GO:0004872; F:receptor activity; IEA.  
DR InterPro; IPR007110; Ig-like.  
KW Receptor.  
KW NON\_TER  
KW NON\_TER  
SQ SEQUENCE 59 AA; 6355 MW; 263284C9F573D36B CRC64;

Query Match 74.4%; Score 29; DB 6; Length 59;  
Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LEHGIN 7  
Db 49 LKHGIN 55

RESULT 19  
Q47667 PRELIMINARY; PRT; 67 AA.  
ID Q47667  
AC Q47667  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE Coli Su+6 (supp) amber suppressor transfer RNA-Leu.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.

```
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85009719; PubMed=6207302;
RA Yoshimura M., Inokuchi H., Ozeki H.;
RT "Identification of Transfer RNA Suppressors in Escherichia coli.";
RL J. Mol. Biol. 177:627-644(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=811893; PubMed=2448476;
RA Nomura T., Fujita N., Ishihama A.;
RT "Expression of the leuX gene in Escherichia coli: regulation at
RT transcription and tRNA processing steps.";
RL J. Mol. Biol. 197:659-670(1987).
DR EMBL; X04174; CAA2774.1; -.
SQ SEQUENCE 67 AA; 7308 MW; 57BBF7985FF84A31 CRC64;

Query Match 74.4%; Score 29; DB 2; Length 67;
Best Local Similarity 83.3%; Pred. No. 1.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 1;

QY 2 EHDGIN 7
DB 48 EYDGIN 53

RESULT 20
O42337 PRELIMINARY; PRT; 68 AA.
ID O42337
AC O42337;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fibroblast growth factor receptor 2b (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Shiozaki C., Asano-Miyoshi M., Tashiro K., Shiokawa K., Emori Y.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D31760; BAA22058.1; -.
DR GO; GO:0004872; Fireceptor activity; IEA.
DR InterPro; IPR007110; IG-like.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 68 68
SQ SEQUENCE 68 AA; 7561 MW; F4AE6489E37B1F07 CRC64;

Query Match 74.4%; Score 29; DB 13; Length 68;
Best Local Similarity 71.4%; Pred. No. 1.5e+02; Mismatches 1; Indels 0; Gaps 0;
Matches 5; Conservative 1;

QY 1 LEHDGIN 7
DB 6 LKHSGIN 12

RESULT 21
Q936F7 PRELIMINARY; PRT; 73 AA.
ID Q936F7
AC Q936F7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=M;
RA Luong T.T., Shu O., Bush K., Lee C.Y.;
RT "The Tyel Capsular Polysaccharide of Staphylococcus aureus is carried
RT in a Staphylococcal Cassette Chromosome Genetic Element.";
EL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U10927; AL26681.1; -.
KW Hypothetical protein.
SQ SEQUENCE 73 AA; 8313 MW; 11766B0808F91B37C CRC64;

Query Match 74.4%; Score 29; DB 2; Length 73;
Best Local Similarity 71.4%; Pred. No. 1.7e+02; Mismatches 1; Indels 0; Gaps 0;
Matches 5; Conservative 1;

QY 1 LEHDGIN 7
DB 50 LEHEAIN 56

RESULT 22
Q8KCM0 PRELIMINARY; PRT; 77 AA.
ID Q8KCM0
AC Q8KCM0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein CT1395.
GN CT1395.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=2210368; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Eison R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL; AE012897; AAM72623.1; -.
DR TIGR; CT1395; -.
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 77 AA; 8578 MW; C9BA18618BEC1356 CRC64;

Query Match 74.4%; Score 29; DB 16; Length 77;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY 1 LEHDG 5
DB 62 LEHDG 66

RESULT 23
Q838W3 PRELIMINARY; PRT; 82 AA.
ID Q838W3
AC Q838W3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN EF0320.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seehadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Seanan M., Nelson W.,
RA Daugherty S., DeSoy R.T., Durkin S., Kolonay J., Madupu R., Khouri H.,
RA Yamathavan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AF016947; AA080183.1; -.
DR TIGR; EF0320; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 82 AA; 9049 MW; BECF70F37AB46120 CRC64;

Query Match 74.4%; Score 29; DB 16; Length 82;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EHDGIN 7
DB 24 EHSIN 29

RESULT 24
Q860C7 PRELIMINARY; PRT; 88 AA.
AC Q860C7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MHC class I antigen (Fragment).
GN ACAR-UA.
OS Acrocephalus arundinaceus (great reed warbler).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.
OX NCBI_TaxID=39621;
RN [1]
SEQUENCE FROM N.A.
RA Westerdaal H., Wittzell H., von Schantz T.;
RT DGGE;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF49697; AAP37625.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00129; MHC_I; 1.
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 10095 MW; B7AFE336C0205727 CRC64;

Query Match 74.4%; Score 29; DB 7; Length 88;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGI 6
DB 56 EHDGI 60

RESULT 25
Q860C6 PRELIMINARY; PRT; 88 AA.
AC Q860C6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MHC class I antigen (Fragment).
GN ACAR-UA.
OS Acrocephalus arundinaceus (great reed warbler).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.
OX NCBI_TaxID=39621;
RN [1]

```

```

OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.
OX NCBI_TaxID=39621;
RN [1]
SEQUENCE FROM N.A.
RA Westerdaal H., Wittzell H., von Schantz T.;
RT DGGE;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF49698; AAP37626.1; -.
DR GO:0016020; C:membrane; IEA.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00129; MHC_I; 1.
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 10101 MW; B575FB54E1C05723 CRC64;

Query Match 74.4%; Score 29; DB 7; Length 88;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGI 6
DB 56 EHDGI 60

RESULT 26
Q860C5 PRELIMINARY; PRT; 88 AA.
AC Q860C5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MHC class I antigen (Fragment).
GN ACAR-UA.
OS Acrocephalus arundinaceus (great reed warbler).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.
OX NCBI_TaxID=39621;
RN [1]
SEQUENCE FROM N.A.
RA Westerdaal H., Wittzell H., von Schantz T.;
RT DGGE;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF49699; AAP37627.1; -.
DR GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00129; MHC_I; 1.
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 10095 MW; B7AFE336C0205727 CRC64;

Query Match 74.4%; Score 29; DB 7; Length 88;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGI 6
DB 56 EHDGI 60

RESULT 27
Q860C4 PRELIMINARY; PRT; 88 AA.
AC Q860C4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MHC class I antigen (Fragment).
GN ACAR-UA.
OS Acrocephalus arundinaceus (great reed warbler).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.
OX NCBI_TaxID=39621;
RN [1]

```

RP SEQUENCE FROM N.A.  
RA Westerdahl H., Wittzell H., von Schantz T.;  
RT "MHC class I typing of great reed warblers by motif-specific PCR and  
RT DGGE";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF449700; AAP37628.1; -;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR001039; MHC\_I.  
DR Pfam; PF00129; MHC\_I; 1.  
DR NON\_TER 1  
FT NON\_TER 88  
SQ SEQUENCE 88 AA; 10095 MW; B7AFE336C0205727 CRC64;  
  
Query Match 74.4%; Score 29; DB 7; Length 88;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 EHDGI 6  
Db 56 EHDGI 60  
  
RESULT 28  
Q860C3 PRELIMINARY; PRT; 88 AA.  
AC Q860C3;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE MHC class I antigen (Fragment).  
GN ACAR-UA.  
OS Acrocephalus arundinaceus (great reed warbler).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.  
OX NCBI\_TaxID=39621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Westerdahl H., Wittzell H., von Schantz T.;  
RT "MHC class I typing of great reed warblers by motif-specific PCR and  
RT DGGE";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF449701; AAP37629.1; -;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR001039; MHC\_I.  
DR Pfam; PF00129; MHC\_I; 1.  
DR NON\_TER 1  
FT NON\_TER 88  
SQ SEQUENCE 88 AA; 10094 MW; BDAF4336C0205727 CRC64;  
  
Query Match 74.4%; Score 29; DB 7; Length 88;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 EHDGI 6  
Db 56 EHDGI 60  
  
RESULT 29  
Q860C9 PRELIMINARY; PRT; 90 AA.  
AC Q860C9;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE MHC class I antigen (Fragment).  
GN ACAR-UA.  
OS Acrocephalus arundinaceus (great reed warbler).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Passeriformes; Sylviidae; Acrocephalus.  
OX NCBI\_TaxID=39621;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Westerdahl H., Wittzell H., von Schantz T.;  
RT "MHC class I typing of great reed warblers by motif-specific PCR and  
RT DGGE";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF449695; AAP37623.1; -;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR001039; MHC\_I.  
DR Pfam; PF00129; MHC\_I; 1.  
DR PRINTS; PR01638; MHCCLASS1.  
DR NON\_TER 1  
FT NON\_TER 90  
SQ SEQUENCE 90 AA; 10644 MW; 7B9E4EF186F5E9D4 CRC64;  
  
Query Match 74.4%; Score 29; DB 7; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 EHDGI 6  
Db 56 EHDGI 60  
  
RESULT 30  
Q8SPLO PRELIMINARY; PRT; 97 AA.  
AC Q8SPLO;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE FGF-7 receptor 2IIIB (Fragment).  
GN FGFR2IIIB.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tissue=Endometrium;  
RA Welter H., Bollwein H., Einspanier R.;  
RT "Expression of horse endometrium";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ439892; CAD29183.1; -;  
DR GO; GO:0004872; F:receptor activity; IEA.  
DR InterPro; IPR007110; Ig-like.  
DR Receptor.  
DR NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 10666 MW; 02C276B3248B7F91 CRC64;  
  
Query Match 74.4%; Score 29; DB 6; Length 97;  
Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LEHDGIN 7  
Db 79 LKHSIN 85  
  
RESULT 31  
Q8SPJ5 PRELIMINARY; PRT; 97 AA.  
AC Q8SPJ5;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE FGF-7 receptor 2IIIB (Fragment).  
GN FGFR2IIIB.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

```
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endometrium;
RA "FGF-system in porcine endometrium.";
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ439895; CAD29186.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
KW Receptor.
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 10666 MW; 02C276B3248B7F91 CRC64;

Query Match 74.4%; Score 29; DB 6; Length 97;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7
DB 79 LKHSGIN 85

RESULT 32
Q9NICO PRELIMINARY; PRT; 98 AA.
AC Q9NICO;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fibroblast growth factor receptor 2 IIb (Fragment).
GN FGFR2IIb.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endometrium;
RA Chen C., Spencer T.E., Bazer F.W.;
RT "Fibroblast growth factor-10: A stromal mediator of epithelial
function in the ovine uterus.";
RL Biol. Reprod. 63:959-966(2000).
DR EMBL; AF213380; AAP26719.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
KW Receptor.
FT NON_TER 1
FT NON_TER 98
SQ SEQUENCE 98 AA; 10767 MW; B372C276B3248B7F CRC64;

Query Match 74.4%; Score 29; DB 6; Length 98;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7
DB 79 LKHSGIN 85

RESULT 33
Q9UMB0 PRELIMINARY; PRT; 100 AA.
AC Q9UMB0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fibroblast growth factor receptor-related protein (Fragment).
GN K-SAM.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=91296403; PubMed=1648704;
RA Champion-Arnaud P., Ronsin C., Gilbert E., Gesnel M.C., Housaint E.,
RA Breathnach R.;
RT "Multiple mRNAs code for proteins related to the BSK fibroblast growth
factor receptor.";
RT Oncogene 6:979-987(1991).
RL factor receptor.
DR EMBL; S40853; AAB19319.1; -.
DR EMBL; S40851; AAB19319.1; JOINED.
DR FIR; A41794; A41794.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR 2fam; PF00047; igf, 1.
DR SMART; SMC0408; IGC2; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Receptor.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 100 AA; 10847 MW; E92E028C34733E34 CRC64;

Query Match 74.4%; Score 29; DB 4; Length 100;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHDGIN 7
DB 50 LKHSGIN 56

RESULT 34
Q8PIF8 PRELIMINARY; PRT; 109 AA.
AC Q8PIF8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Benzene 1,2-dioxygenase ferredoxin protein.
GN Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Perro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., Monteiro-Vitorello C.B., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan P., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyana A.M., Kishi L.T., Leite R.F., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.N., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:455-463(2002).
DR EMBL; AF011936; AAM37785.1; -.
DR GO; GO:0016702; F:oxidoreductase activity, acting on single d. . .; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR005806; Rieske_dom.
```

DR Pfam; PF00355; Rieske; 1.  
 KW Dioxigenase; Complete proteome.  
 SQ SEQUENCE 109 AA; 12250 MW; 04024C66C2B5EA23 CRC64;  
 Query Match 74.4%; Score 29; DB 16; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGI 6  
 DB 99 EHDGI 103  
 RESULT 35  
 Q8W6N9 PRELIMINARY; PRT; 117 AA.  
 AC Q8W6N9;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Gp61.  
 GN 61.  
 OS Bacteriophage phiE125.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
 OC Lambda-like viruses.  
 OC NCBI\_TaxID=180504;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Woods D.E., Jeddeloh J.A., Fritz D.F., DeShazer D.;  
 RT "Burkholderia thailandensis E125 Harbors a Temperate Bacteriophage  
 Specific for Burkholderia mallei";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AF447491; AAL40335.1; -  
 SQ SEQUENCE 117 AA; 13882 MW; BF8C2A4B3B1D82BE CRC64;  
 Query Match 74.4%; Score 29; DB 9; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 EHDGI 6  
 DB 14 EHDGI 18  
 RESULT 36  
 Q8PE20 PRELIMINARY; PRT; 117 AA.  
 AC Q8PE20;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein XCC0163.  
 GN XCC0163.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OC NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=ATCC 33913 / NCPPB 528;  
 RC MEDLINE=22022145; PubMed=12024217;  
 RX da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the Genomes of two Xanthomonas pathogens with differing  
 host specificities";  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE012112; AM39482.1; -  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; ehand; 2.  
 DR PROSITE; PS00018; EF\_HAND; 3.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 117 AA; 12918 MW; DE9063AC8DF999B6 CRC64;  
 Query Match 74.4%; Score 29; DB 16; Length 117;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LEHDGI 6  
 DB 64 LEHDGI 69  
 RESULT 37  
 Q8KSU0 PRELIMINARY; PRT; 124 AA.  
 AC Q8KSU0;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical protein.  
 OS Rhizobium leguminosarum (biovar viciae).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OC NCBI\_TaxID=367;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22151066; PubMed=12019272;  
 RX Basu S.S., Karbarz M.J., Raetz C.R.H.;  
 RT "Expression Cloning and Characterization of the C28 Acyltransferase of  
 Lipid A Biosynthesis in Rhizobium leguminosarum";  
 RL J. Biol. Chem. 277:28959-28971(2002).  
 DR EMBL; AF510733; AA042955.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 124 AA; 13113 MW; 390E5F95F3CFB5E5 CRC64;  
 Query Match 74.4%; Score 29; DB 2; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LEHDGI 5  
 DB 98 LEHDGI 102  
 RESULT 38  
 Q9HBN2 PRELIMINARY; PRT; 137 AA.  
 AC Q9HBN2;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,  
 RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,  
 RA Yu J., Han L.H.;  
 RT "Novel Human cDNA clones with function of inhibiting cancer cell



RT growth.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF218028; AAG17270.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 137 AA; 15826 MW; 90866F1966195687 CRC64;

Query Match 74.4%; Score 29; DB 4; Length 137;  
Best Local Similarity 57.1%; Pred.No. 3.3e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHGDIN 7  
|:|:|  
Db 109 LQHGVN 115

## RESULT 39

QSDH14 PRELIMINARY; PRT; 138 AA.  
ID Q9DH14;  
AC Q9DH14;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 129R protein.  
GN 129R.  
OS Yaba-like disease virus (YLDV).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Yatapoxvirus.  
OX NCBI\_TaxID=132475;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lee H.J.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21176366; PubMed=11277691;  
RA Lee H.J., Essani K., Smith G.L.;  
RT "The genome sequence of Yaba-like disease virus, a yatapoxvirus.";  
RL Virology 281:170-192(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Lee H.J.;  
RL Thesis (2000), Sir William Dunn School of Pathology, University of.  
DR EMBL; AJ293568; CAC21367.1; -.  
SQ SEQUENCE 138 AA; 16059 MW; 100FA621BDAB4108 CRC64;

Query Match 74.4%; Score 29; DB 12; Length 138;  
Best Local Similarity 85.7%; Pred.No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEHGDIN 7  
|:|:|  
Db 3 LEHDIIN 9

## RESULT 40

Q9Y1T3 PRELIMINARY; PRT; 141 AA.  
ID Q9Y1T3;  
AC Q9Y1T3;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Metallothionein-like protein (fragment).  
OS Jasus edwardsii (red rock lobster).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;  
OC Palinuroidea; Palinuridae; Jasus.  
OX NCBI\_TaxID=95461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Khoo J.G.I., Sin F.Y.T.;  
RT "Novel peptide gene sequence from the lobster Jasus edwardsii.";  
RL Zool. Stud. 38:95-109(1999).  
DR EMBL; AF091369; AAD37832.1; -.

FT NON TER 1 1  
SQ SEQUENCE 141 AA; 14893 MW; 09551B769BB01430 CRC64;  
Query Match 74.4%; Score 29; DB 5; Length 141;  
Best Local Similarity 100.0%; Pred.No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|:|:|  
Db 36 LEHDG 40

## RESULT 41

Q97XH7 PRELIMINARY; PRT; 146 AA.  
ID Q97XH7;  
AC Q97XH7;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein SSO1764.  
GN SSO1764.  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=2287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=1142726;  
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
Rawayez M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,  
De Moers A., Brauso G., Fletcher C., Gordon P.M.K.,  
Ra Heikamp-de Jong I., Jeffries A.C., Koza C.J., Medina N., Peng X.,  
Ra Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
Ra Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
Ra Garrett R.A., Ragan M.A., Sersen C.W., Van der Oost J.;  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
EL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
DR EMBL; AS006787; AAK41957.1; -.  
DR PIR; F90337; F90337.  
DR InterPro; IPR005544; HHE.  
DR Pfam; PF03794; HHE; 2.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 146 AA; 17058 MW; 039AA714C98B5E8D CRC64;

Query Match 74.4%; Score 29; DB 17; Length 146;  
Best Local Similarity 100.0%; Pred.No. 3.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
|:|:|  
Db 101 LEHDG 105

## RESULT 42

Q972G0 PRELIMINARY; PRT; 146 AA.  
ID Q972G0;  
AC Q972G0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein ST1173.  
GN ST1173.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX MEDLINE=21456156; PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,

RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermoacidophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL; AF000985; BAB66208.1; -;  
 DR InterPro; IPR005544; HHE.  
 DR Pfam; PF03794; HHE; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 146 AA; 17266 MW; EA2A859AF0485D94 CRC64;

Query Match 74.4%; Score 29; DB 17; Length 146;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEHDG 5  
 |||||  
 DB 101 LEHDG 105

## RESULT 43

Q89FE7 PRELIMINARY; PRT; 147 AA.  
 AC Q89FE7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE BL-6753 protein.  
 GN BLR6753.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 OX NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Idegawa K., Iriiguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110.";  
 RL DNA Res. 9:193-197(2002).  
 DR EMBL; AF005959; BAC52018.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 147 AA; 15940 MW; 68C38FE059012BB CRC64;

Query Match 74.4%; Score 29; DB 16; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGI 6  
 |||||  
 DB 66 EHDGI 70

## RESULT 44

O34195 PRELIMINARY; PRT; 22 AA.  
 AC O34195;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE GROES (Fragment).  
 GN GROES.  
 OS Ehrlichia risticii.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Anaplasmataceae; Neorickettsia.  
 OX NCBI\_TaxID=950;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=97373904; PubMed=9230387;  
 RA Sumner J.W., Nicholson W.L., Maessung R.F.;  
 RT "PCR amplification and comparison of nucleotide sequences from the  
 RT groESL heat shock operon of Ehrlichia species.";  
 RL J. Clin. Microbiol. 35:2087-2092(1997).  
 DR EMBL; U96732; AAB65632.1; -;  
 DR InterPro; IPR001476; Chaprinin\_Opn10.  
 DR Pfam; PF00166; cpn10; 1.  
 FT NON TER 1  
 SQ SEQUENCE 22 AA; 2504 MW; 21D959D75191F620 CRC64;

Query Match 71.8%; Score 28; DB 2; Length 22;  
 Best Local Similarity 80.0%; Pred. No. 70;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 EHDGI 6  
 |||||  
 DB 1 EHDGI 5

## RESULT 45

Q81RR9 PRELIMINARY; PRT; 69 AA.  
 AC Q81RR9;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN BA1971.  
 OS Bacillus anthracis (strain Ames).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=198094;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608414; PubMed=12721629;  
 RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
 RA Nelson K.E., Tectelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.F., Okstad O.A., Helgason E., Ristone J., Wu M.,  
 RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
 RA DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benson J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
 RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
 RA Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86(2003).  
 DR EMBL; AE017030; AAP25863.1; -;  
 DR TIGR; BA1971; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 69 AA; 7803 MW; 687CAF89236197C9 CRC64;

Query Match 71.8%; Score 28; DB 16; Length 69;  
 Best Local Similarity 80.0%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HDGIN 7  
 |||||  
 DB 27 HDGIN 31

Search completed: June 15, 2004, 12:24:12  
 Job time : 41 secs